

Access DB# 39046

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched.
 Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or
 utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if
 known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: AB5501
 Searcher Phone #: 308-4501
 Searcher Location: Biotek Lab
 Date Searcher Picked Up: 3/30/01
 Date Completed: 4/4/01
 Searcher Prep & Review Time: _____
 Clerical Prep Time: 3 mi
 Online Time: 2 mi

Type of Search

NA Sequence (#) 1
 AA Sequence (#) _____
 Structure (#) _____
 Bibliographic _____
 Litigation _____
 Fulltext _____
 Patent Family _____
 Other _____

Vendors and cost where applicable

STN _____
 Dialog _____
 Questel/Orbit _____
 Dr.Link _____
 Lexis/Nexis _____
 Sequence Systems AB5501
 WWW/Internet _____
 Other (specify) _____

Result No.	Score	Query Match	Length	DB	ID	Description
1	1380	100.0	1380	20	X88501	A. gossypii oroid
2	370.4	26.8	2330	17	W08603	UA3 gene. Candid
3	370.4	26.8	2330	19	V15105	Base sequence of E
4	369.6	26.8	1469	15	O66476	K. maritimus UA3
5	357	25.9	5181	20	X85877	Nucleotide sequenc
6	357	25.9	7063	20	X85878	Synthetic operon c
7	356	25.8	1101	20	V83956	Bacterial artificial
8	356	25.8	4102	19	V64257	Plasmid pPK9/10 DN
9	356	25.8	4164	16	T04575	Plasmid pAT-1 sequ
10	356	25.8	4164	19	V22271	pAT-1 (pSD544). S
11	356	25.8	4164	20	V69740	Nucleotide sequenc
12	356	25.8	4232	20	X60602	S. cerevisiae ATH1

[illegible]

FT CDS 1259..2062
 FT /*tag= a
 FT /product= URA3 gene product.
 XX
 XX W09532289-A1.
 XX
 XX 30-NOV-1995.
 PD
 XX
 XX 25-MAY-1995; 95MO-JP01005.
 PF
 XX
 PR 28-APR-1995; 95JP-0129287.
 PR 25-MAY-1994; 94JP-0135015.
 PR 26-OCT-1994; 94JP-0285823.
 PA
 XX
 PA (KIRI) KIRIN BEER KK.
 PA (KIRI) KIRIN BREWERY KK.
 PI
 XX
 PI Kojima S, Kondo K, Mitsuwa N:
 DR WPI; 1996-020584/02.
 DR P-PSDB; R77659.
 XX
 XX Ribosomal protein L41 gene, promoter and terminator sequences,
 PT isolated from Candida utilis - used in vector for expression of
 PT hetero:gene(s) in yeast
 XX
 PS Claim 12; Page 153-156; 252pp; Japanese.
 XX
 CC The L41 ribosomal structural protein gene (708602) can be used in the
 CC construction of expression vectors designed specifically to express
 CC heterologues in yeast. The vectors also comprise: a promoter and
 CC terminator sequence selected from phosphoglyceric acid kinase (Pck),
 CC glyceraldehyde-3-phosphoric acid dehydrogenase (GAP) and
 CC protoplasmic membrane proton ATPase (PMA) genes isolated from
 CC Candida utilis; a drug resistance marker selected from
 CC aminoglycoside-3'-phosphotransferase or hygromycin B
 CC phosphotransferase; and the heterogene to be expressed. L41 confers
 CC cyclohexamine resistance on the yeast. The expression vector may
 CC also contain C. utilis RNA gene cluster sequences,
 CC orchidin-5'-phosphate decarboxylase, or URA3 gene sequences.
 CC
 SQ Sequence 2330 BP; 669 A; 498 C; 613 G; 550 T; 0 other;

Query Match 26.8%; Score 370.4; DB 17; Length 2330;
 Best Local Similarity 67.5%; Pred. No. 1.9e-96;
 Matches 521; Conservative 0; Mismatches 251; Indels 0; Gaps 0;

Oy 214 caagcaaatcttcgcaagaagcgaagcacaacatctgcagcttcgcaagaagcttc 273
 Db 1266 ccaagcaaatcttcgcaagaagcgaagcacaacatctgcagcttcgcaagaagcttc 1325
 Oy 274 tggcatgcatgcaagaagaacaacatctgcagcttcgcaagaagcttcgcaagaagcttc 333
 Db 1326 ttcgccttaccgagtgccaagaagaacacgactgctgctcagtgctgctgctgcaacag 1385
 Oy 334 gaaagctctcggagctgcaagaacgctggaacgacacattgtctgctgcaagaacagcttc 393
 Db 1386 aggaagtgctcgaagctgctgcaagaacgctggaacgacacattgtctgctgcaagaacagcttc 1445
 Oy 394 tgcgaacacgcaagcttcgcaagaacgctggaacgacacattgtctgctgcaagaacagcttc 453
 Db 1446 tgcgaacacgcaagcttcgcaagaacgctggaacgacacattgtctgctgcaagaacagcttc 1505
 Oy 454 ctaagcaacacgcaagcttcgcaagaacgctggaacgacacattgtctgctgcaagaacagcttc 513
 Db 1506 aggaagtgctcgaagcttcgcaagaacgctggaacgacacattgtctgctgcaagaacagcttc 1565
 Oy 514 agctgcagtgctcctcctcggcggtgctacgctgctgcaagaagtggaagcgatataacacagcttc 573
 Db 1566 aggaagtgctcctcctcggcggtgctacgctgctgcaagaagtggaagcgatataacacagcttc 1625
 Oy 574 agcgcgctcaccgagcccgcggtgctacgctgctgcaagaagtggaagcgatataacacagcttc 633

Db 1626 accgctgcaacgctgcaagtgatcgcgaagaaggttgcaagaagctgcacagaagaacacagc 1685
 Oy 634 aggaacacgaggggtgctgcaagtgatcgcgaagaagcttcctcgaagaagcttcgcaagc 693
 Db 1686 atgagccaaagagggctgctgcaagtgatcgcgaagaagcttcctcgaagaagcttcgcaagc 1745
 Oy 694 gaaacataacccgagggctgctgcaagtgatcgcgaagaagcttcctcgaagaagcttcgcaagc 753
 Db 1746 gaaacataacccgagggctgctgcaagtgatcgcgaagaagcttcctcgaagaagcttcgcaagc 1805
 Oy 754 tcatcgcaagcagctgcaagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagc 813
 Db 1806 tcatcgcaagcagctgcaagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagc 1865
 Oy 814 cgggggtgctgctgcaagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagc 873
 Db 1866 cgggggtgctgctgcaagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagc 1925
 Oy 874 agctgcagcagcagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagaagc 933
 Db 1926 agctgcagcagcagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagaagc 1985
 Oy 934 ggaaccccaagtgcaagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagc 993
 Db 1986 ggaaccccaagtgcaagtgatcgcgaagaagcttcgcaagaagcttcgcaagaagcttcgcaagc 2037

RESULT 3
 V31505
 ID V31505 standard; DNA; 2330 BP.
 AC V31505;
 XX
 DT 14-AUG-1998 (first entry)
 DE Base sequence of DNA fragment containing yeast URA3 gene.
 XX
 KW Candida utilis; yeast vector; promoter; marker gene; GAP gene;
 KW high-efficiency integration; monellin; food; drug; L41; URA3; ss.
 XX
 OS Candida utilis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1259..2062
 FT /*tag= a
 FT /note= "URA3 gene segment"
 XX
 PN W09807873-A1.
 PD 26-FEB-1998.
 XX
 PF 22-AUG-1997; 97MO-JP02924.
 PR 23-AUG-1996; 96JP-0241062.
 PA (KIRI) KIRIN BEER KK.
 PA Kondo K, Miura Y;
 PI
 XX
 DR WPI; 1998-169177/15.
 DR P-PSDB; W57489.
 XX
 PT Yeast vector for multi-copying on to chromosomes of yeast such as
 PT Candida utilis - contains a shortened promoter sequence linked to a
 PT marker gene for high-efficiency integration
 PS Claim 28; Fig 5; 107pp; Japanese.
 CC This represents the base sequence of DNA fragment containing yeast URA3
 CC gene. This can be used in the construction of a yeast vector for
 CC multi-copying on to chromosomes of yeast such as Candida utilis. The
 CC vector contains a DNA homologous with a chromosomal gene of the yeast

CC (preferably ribosomal DNA (rDNA) such as the URA3, LAI, PGK, GAP or PMA
 CC gene), a marker gene to be used in transformant selection (such as a
 CC drug resistance gene, e.g. the cycloneximide resistance gene LAI, the
 CC 6418 resistance gene Tr903-APR, or the hygromycin B resistance gene
 CC (from E.coli) HPT), a shortened promoter sequence such as the C.utillis
 CC LAI, phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate
 CC dihydrogenase (GAP) or plasma membrane proton ATPase (PMA) gene promoter
 CC and a gene of interest from yeast or other origin. The vectors are useful
 CC in the preparation of proteins for food or drug use in high efficiency.
 CC They can be used especially for the production of single-chain monellin,
 CC which is a low-calorie sweetener whose thermostability is greater than
 CC that of the dimeric natural monellin. Multiple copies of the desired gene
 CC are integrated into the yeast chromosome and high expression efficiency
 CC is obtained.

SQ Sequence 2330 BP; 669 A; 498 C; 613 G; 550 T; 0 other:

Query Match 26.8%; Score 370.4; DB 19; Length 2330;
 Best Local Similarity 67.5%; Pred. No. 1,9e-96;
 Matches 521; Conservative 0; Mismatches 251; Indels 0; Gaps 0;

QY 214 caacgaatctctacgagaagaagggccaagtcacacatctgcagctgctagaagcttc 273
 DB 1266 ccagctatctgacacaggaagggacatcgacagcccttcgcactgttaagcttct 1325
 QY 274 tggcatctgacaggaagaacacatctctgctcctctatgtgaggaagctcta 333
 DB 1326 tctcgctatctgagatcagaagaagacacctctgtgacagctgctgacacag 1385
 QY 334 gaagatctctgagctagacagacagcagctgcagacacattgctgcaggaacatct 393
 DB 1386 aggaagcttgcagacagcagctgctgacagctgctcctatctcgtctgttgaagaagcata 1445
 QY 394 tgcacatactgacagacatctgcacatcgacagacagctgcagacagcttcgag 453
 DB 1446 tctgatactgagctctctcctacagctgagctgacagctgctgagcttca 1505
 QY 454 ctaagacacatctcagatctctcgaagacagctgctgacatctgacacagctta 513
 DB 1506 agaaagacacatctcctcctctctgagacagctgaagcttgcagatctgcgacacagctta 1565
 QY 514 agctgcagtaactctcctcgcgctgacagctgcagagctgcagatctgacacagctta 573
 DB 1566 aggcacagatcagcgcgctgctgacagatctgcgacagctgcagatctgacacagctta 1625
 QY 574 acggcgtcacccgcgctgctgacagcgcgctgcagagctgcagacagctgcacac 633
 DB 1626 acggctgcacccgcgctgacagctgcagagctgcagagctgcagacagctgcacacag 1685
 QY 634 aggaacccagggggtgctgacagctgcagagctgcagagctgcagagctgcagagctgc 693
 DB 1686 atgagccaaagagggctgctgacagctgcagagctgcagagctgcagagctgcagagctgc 1745
 QY 694 gaagacataccgcgagctgcctgacagctgcagagctgcagagctgcagagctgcagagctgc 753
 DB 1746 ggcacataccgcgagagctgcctgacagctgcagagctgcagagctgcagagctgcagagctgc 1805
 QY 754 tcatcgacagctgcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 813
 DB 1806 tcatcgacagagagagctgcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 1865
 QY 814 cgggggttggcctgacagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 873
 DB 1866 caggcgttggcctgacagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 1925
 QY 874 agatctcagcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 933
 DB 1926 agatctcagcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 1985
 QY 934 gaagaccacagctgcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 985
 DB 1986 gaagatcacaacagctgcagagctgcagagctgcagagctgcagagctgcagagctgcagagctgc 2037

RESULT 4
 ID 066476 standard; DNA; 1469 BP.
 AC 066476;
 DT 17-FEB-1995 (first entry)
 DE K. marxianus URA3 gene.
 KW Polymerase chain reaction; primer; amplify; PCR; extension assay;
 KW upstream activating sequence; upstream repressing sequence; UAS;
 KW terminator; secretory signal; secretion; yeast; transformation;
 KW protein; metabolite; ss.
 OS Kluyveromyces marxianus.
 FH Key Location/Qualifiers
 FT CDS 546..1349
 FT /*tag= a

W09413821-A.
 PD 23-JUN-1994.
 PF 09-DEC-1993; 93WO-EP03547.
 PR 11-DEC-1992; 92EP-0203932.
 PR 29-JAN-1993; 93EP-0200240.
 PA (UNIL) QUEST INT BV.
 PI Chapman JW, Musters W, Rouwenhorst RJ, Toschka HY, Verbakel JMA;
 DR WPI: 1994-217894/26.
 DR P-PSDB: R56259.

PT Yeast regulatory sequence from the inulinase gene - for protein,
 RNA or metabolite prodn.
 PS Example 13; Fig 12; 68pp; English.
 XX

CC This sequence represents URA3 gene from K. marxianus. This gene was
 CC used in an example in the construction of the uracil requiring K.
 CC anus mutants, KMS3. This gene is involved in the regulation of the
 CC biosynthesis of pyrimidine. Yeast cells which have an active URA3
 CC gene are unable to grow in medium containing 5-fluoro-orotic acid
 CC (5-FOA). KMS3 is a non-reverting K. marxianus leu2 strain which is
 CC suitable for overexpression of homologous or heterologous proteins in
 CC the food industry.
 CC XX

SQ Sequence 1469 BP; 443 A; 258 C; 356 G; 412 T; 0 other:

Query Match 26.8%; Score 369.6; DB 15; Length 1469;
 Best Local Similarity 66.1%; Pred. No. 2.6e-96;
 Matches 534; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

QY 194 acctagcatgacacatgctacagaaatcttaccgagaagggccaagtcacacatctc 233
 DB 530 aactaatcaagaagagatgctgactaagtgatctactcgagaagcagctgctcatagaag 589
 QY 254 gccagttgctagaagcttctgcatgtgacgacgagaagaacaacaaatctctgcgcttc 313
 DB 590 tccagttgctcgaactttaaacttgatggaagaagaagtcacaaactaagtgttc 649
 QY 314 ccttgatgtgagcagctctgaagaagctctggaagctgacgacagctggaacgacat 373
 DB 650 tcttgatgtgctgaataaagcagagctgctaaagatgctgaggttcttggtccatataat 709

OY	374	ctgtctctggaagacacattccacatactcgaagagcttcgaatctgagacagctca	433
Db	710	ctgtctatctggaagacacattgtagatattcttgaggatcttcagctttggaataacattgt	769
OY	434	gctgcctcagcagctctgctgcctaaagcaattcatgatcttcgagagccgaagcttcgc	493
Db	770	gctccttgaaagacattaaacgaagaaacacaaagctttcttgatatattgaaagacaagaaatttcg	829
OY	494	tgacatttgcaacacacggttttaagctgcagtagtctctccgcgcgttgtaacgfatcgcggagtg	553
Db	830	cgacatttgggaacacccctttaaattacaatacacgctctgtgtgtatccagcatalccgcgaag	889
OY	554	ggcggatatttcaacaaatgcacacggtgcataccggtcccccgttggtatagcccggttcgaaga	613
Db	890	gtctgatatcccaacattgcacacggtgtgagatctgtgcgggcatgtgtctgtcttgaaagca	949
OY	614	ggctgcgaaacttgccctcacagaagaaacccaaagggtctgtcgtatgcttgagcagaagctctctc	673
Db	950	aggtgtccgaagaaagtctcagaagaaacccaaagggtctgtcgtatgcttgagcagaagctctctc	1009
OY	674	tcaaggctctcttgccgcgcgcggagaactaaaccgcgggcgtctcgtatgaaatgtgcgaagcttga	733
Db	1010	caaaagggtctctcgaacggttgataacactcgttgaggaccgttggaatattgccaaagtgtga	1065
OY	734	cgaagaactttgtgatctgggtttcatctgcgcagcgttgacaaatgggttgggcgtgtgcagaacgctt	793
Db	1070	taaggaaacttcttatcttggaatttatctgcacaaacgatactgtgtgaaagaaagaaagaggtcta	1125
OY	794	tgactctgtctcatcatctgaccccggggggtcttgacccctgcagacagaagaagagagcgtctgcga	853
Db	1130	cgattgttgtgatactgacgcgcgaagggtgtgtgtcttgatgagcaaaagtatgactcttgagaca	1188
OY	854	gcaatataccgcacggttggaatgaggtctgctcagcgaaggttaacgaatgtgatactgtttgcgaag	913
Db	1190	acaatacaagaacattgtgataagattctgtccgggttgagatacagaatattatattgtgtgag	1249
OY	914	agggctctcttgcacaaagggaagagaccccaagggtcgaagggttgcccgcttaccgcgaagcgcg	973
Db	1250	aggtctcttgcgaagaagggaagagatactctgtatgtgaaaggttgaaagatacagaagaagcgagg	1305
OY	974	ttgggaagcttaactcttcgcgcgctatgggc	1001
Db	1310	atggagacgtcttaacttgaaagagatgaagc	1337
RESULT 5			
X85877/c			
ID	X85877 standard; DNA; 5181 BP.		
XX	X85877;		
AC			
XX			
DT	13-SEP-1999 (first entry)		
XX			
DE	Nucleotide sequence of plasmid YEp352.		
XX			
KW	Methylglyoxal synthase; MGS; ethanol production; 1,2-propanediol;		
KW	unsaturated polyester resin; liquid laundry detergent; cosmetic;		
KW	antifreeze; deicing formulation; carbon dioxide; alcohols;		
KW	organic acid; animal feed; ss.		
XX			
OS	Synthetic.		
XX			
PN	WO9928481-A1.		
XX			
PD	10-JUN-1999.		
XX			
PF	30-NOV-1998; 98WO-US25318.		
XX			
PR	03-DEC-1997; 97US-0984717.		
XX			
PA	(WISC) WISCONSIN ALUMNI RES FOUNDD.		
XX			
XX	Cameron DC, Hoffman ML, Shaw AJ;		
XX			

CC PAT-2 (PSD545) includes: the yeast URA3 gene; a bacterial origin of
 CC replication and multicloning site; a selectable marker (DHFR) gene; 2
 CC blunt-ended transposon termini (substrates for yeast retrotransposon
 CC Tyl integrase) which flank a restriction site useful for insertion of
 CC a second selectable marker gene; and an XmnI site that flanks the
 CC transposon terminus. Digestion of PAT-2 at the XmnI site liberates
 CC a primer island artificial transposon (PATr, see T04570) useful for
 CC DNA sequencing and genome mapping.

XX Sequence 4933 BP: 1431 A; 1031 C; 1157 G; 1314 T; 0 other:

Query Match 25.8%; Score 356; DB 16; Length 4933;
 Best Local Similarity 65.5%; Pred. No. 3,7e-92;
 Matches 521; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

```

OY 209 catgtcaagaatacttcacgaagaaggcacaacattccagctgtctagaaa 268
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 416 catgtcgaagaatacttcacgaagaaggcacaacattccagctgtctagaaa 475
OY 269 gcttcgcatgtgacgacgagagaagaacacatctcgccttccttgatgagac 328
    || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 476 gctatttaatacagcagaagaacacacaaacttgctgtcattcattgattcgtac 535
OY 329 gctcgaagaagctctcgaagctagacagacacgctgagacgacatttgcctgaagac 388
    | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 536 caccagaagaattactgaggttaggttagaagcattagctcccaaatcttctacataaac 595
OY 389 acattgcgaatactgacgaggaacttcgacatcgaagacagatcagcgcgtgcagcagt 448
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 566 acattgcgaatactgacgaggaacttcgacatcgaagacagatcagcgcgtgcagcagt 655
OY 449 tgcgcgaagacacaaacttcacgaagaaggcacaacattcctgacattgcaaacac 508
    | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 656 atccgcgaagacacaaacttcacgaagaaggcacaacattcctgacattgcaaacac 715
OY 509 gcttaagctcgaactcctccgcgcgtgtagacgatacgcgagtgagcgatataccaa 568
    | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 716 agtcaaatcgaactcctcgcgcgtgtagacgatacgcgagtgagcgatataccaa 775
OY 569 tgcacaagcgctcgaacggcccgctgtagatagcgcgcgtgtagatagcgcgcgtgtagat 628
    ||||| | | | | | | | | | | | | | | | | | | | | | |
DB 776 tgcacaagcgctcgaacggcccgctgtagatagcgcgcgtgtagatagcgcgcgtgtagat 835
OY 629 ctcaagaagaacccagcggtgctgtagatagcgcgcgtgtagatagcgcgcgtgtagat 688
    || | | | | | | | | | | | | | | | | | | | | | | | |
DB 836 aacaagaagaacccagcggtgctgtagatagcgcgcgtgtagatagcgcgcgtgtagat 895
OY 689 ggcgggaagaatacttcacgaagaaggcacaacattcctgacattgcaaacac 895
    ||||| | | | | | | | | | | | | | | | | | | | | | |
DB 896 taccggaagaatacttcacgaagaaggcacaacattcctgacattgcaaacac 955
OY 749 cgggttcacgcgacgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 808
    || | | | | | | | | | | | | | | | | | | | | | | | |
DB 956 cgggttcacgcgacgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 1015
OY 809 gaacccggggttgccttgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 868
    || | | | | | | | | | | | | | | | | | | | | | | | |
DB 1016 gaacccggggttgccttgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 1075
OY 869 ggaagagctcgcgacgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 928
    ||||| | | | | | | | | | | | | | | | | | | | | | |
DB 1076 ggaagagctcgcgacgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 1135
OY 929 ggaagagctcgcgacgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 988
    ||||| | | | | | | | | | | | | | | | | | | | | | |
DB 1136 ggaagagctcgcgacgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 1195
OY 989 ggcgcgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 1004
    | | | | | | | | | | | | | | | | | | | | | | | |
DB 1196 ggaagagctcgcgacgctgtagatagcgcgcgtgtagatagcgcgcgtgtagatagcgcgcgt 1211

```

RESULT 14

V22272
 ID V22272 standard; DNA; 4933 BP.
 AC V22272;
 XX 07-JUL-1998 (first entry)

XX PAT-2 (PSD545).

XX Target DNA sequence determination; PAT-2; PSD545; circular;
 KW cyclic; ds.

XX Synthetic.

PN US5728551-A.

PD 17-MAR-1998.

PF 02-MAR-1995; 95US-0397679.

PR 02-MAR-1995; 95US-0397679.

PR 02-MAR-1994; 94US-0204675.

XX (UYJO) UNIV JOHNS HOPKINS.

XX Boeke JD, Bratlerman LT, Devine SE;
 WPI; 1998-270435/24.

PT DNA sequencing - using artificial transposon(s) inserted into
 PT target DNA by incubation with retroviral or retrotransposon
 PT integrase in vitro

XX Example; Columns 27-32; 38pp; English.

The present sequence was used in the development of a novel method
 for the determination of a target DNA sequence. The method
 comprises: (a) incubating target DNA with a retroviral or
 retrotransposon integrase and an artificial transposon whose
 termini are substrates for the integrase, to form target DNA
 molecules with quasi-randomly integrated insertions of the
 artificial transposon; (b) transforming host cells with the
 artificial transposon-containing target DNA molecules; (c) selecting
 transformants harbouring transposon-containing target DNA;
 for use as a DNA sequencing template; (e) hybridising a primer to
 the isolated transposon-containing target DNA, the primer being
 complementary to a portion of the artificial transposon; and (f)
 extending the primer to determine a nucleotide sequence flanking
 the artificial transposon. The integrase is yeast retrotransposon
 Tyl integrase. The target DNA is a plasmid or cosmid. The integrase
 is supplied as Tyl virus-like particles. The termini contain Tyl U3
 sequences. The termini consist of 4-11 bp. The artificial
 transposon is provided by digestion with a restriction enzyme that
 generates A/C blunt ends, especially XmnI. Step (b) is effected by
 electroporation. The molar ratio of artificial transposon to target
 DNA is at least 2.5:1. Target DNA can be engineered to convert
 virtually any DNA sequence or combination of sequences into an
 artificial transposon, so custom transposons having any desired
 feature can be easily designed and constructed.

XX Sequence 4933 BP: 1431 A; 1031 C; 1157 G; 1314 T; 0 other:

Query Match 25.8%; Score 356; DB 16; Length 4933;
 Best Local Similarity 65.5%; Pred. No. 3,7e-92;
 Matches 521; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

```

OY 209 catgtcaagaatacttcacgaagaaggcacaacattccagctgtctagaaa 268
    ||||| | | | | | | | | | | | | | | | | | | | | | | | |
DB 416 catgtcgaagaatacttcacgaagaaggcacaacattccagctgtctagaaa 475

```

QY 269 gctctgacatgacagagagaagaacacatctcgtcgtcccttgatgctggac 328
 DB 476 gctatttaatatatgacgaagaagaacaaacttgctgctcattggatctgac 535
 QY 329 gctagaagaactctgagagtagacagacacgctggagccgacatctgctgtagaac 388
 DB 536 caccagaagaaatctcgtgagtagatgaaagcatagatcccaaaattgcttaataaac 555
 QY 389 acatgctgacatactgacggaacttcgacatcgagacagacatcaagccgctgacagact 448
 DB 596 acatgtagatactctgactgattcttcacatggagggacagcttaagccgctaaagcatt 655
 QY 449 tgggagctaaagacacactcgtcttcctgagagccgacagctgctgacttgacacac 508
 DB 656 atccgcaagtaacaaatttctactctcgagacagaacaaattgctgacttgatgac 715
 QY 509 ggttaagctgacatctcctcgcgctgacatccgctatccgagtgaggatattaccaa 568
 DB 716 agtcaaaatgacgactcctgctgctgctgctatcagaaatagacgaatggcgacacttcgaa 775
 QY 569 tgcacacgctgacacgagcccgctgtagacgagctgaaagagctgacgaactggc 628
 DB 776 tgcacacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 835
 QY 629 ctcaagaggaacccagaggggtgctgctgctgctgctgctgctgctgctgctgctgct 688
 DB 836 aacaaaggaacccagagggcttctgctgctgctgctgctgctgctgctgctgctgctgct 895
 QY 689 gcgcgaggaacataccgagcgctgctgctgctgctgctgctgctgctgctgctgctgct 748
 DB 896 tactcgaggaatatacctgacgctgctgctgctgctgctgctgctgctgctgctgctgct 955
 QY 749 cgggttcatctgacgagctgacatgctgctgctgctgctgctgctgctgctgctgctgct 808
 DB 956 cggcttcatctgacgaggaacacatgctgctgctgctgctgctgctgctgctgctgctgct 1015
 QY 809 gaccgagggctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 868
 DB 1016 gacacccggctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1075
 QY 869 ggaatgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 928
 DB 1076 ggaatgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1135
 QY 929 ggaagaggaacccagagctgacatgctgctgctgctgctgctgctgctgctgctgctgct 988
 DB 1136 ggaagaggaacccagagctgacatgctgctgctgctgctgctgctgctgctgctgctgct 1195
 QY 989 ggcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1004
 DB 1196 ggcgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1211
 RESULT 15
 V69741
 ID V69741 standard; DNA: 4933 BP.
 AC V69741:
 DT 04-FEB-1999 (first entry)
 DE Nucleotide sequence of artificial transposon AT-2 (part-2) plasmid.
 XX
 XX Transposon; artificial; marker gene; U3; yeast; retrotransposon;
 KW Tyl; DNA mapping; DNA sequencing; primer site; part-2; ds.
 XX
 OS Synthetic.
 OS Saccharomyces cerevisiae.
 XX
 XX US5843772-A.
 PN 01-DEC-1998.
 XX

PF 10-JUN-1996; 96US-0660754.
 XX 02-MAR-1995; 95US-0397679.
 PR 02-MAR-1994; 94US-0204675.
 PR 10-JUN-1996; 96US-0660754.
 XX
 XX (U3) UNIV JOHNS HOPKINS.
 XX
 PI Boeke JD, Bralerman LT, Devine SE;
 DR WPI: 1999-044590/04.
 XX
 PT Artificial transposons with terminal Tyl U3 sequences - useful for
 XX DNA sequencing
 XX
 PS Examples; Fig 9A-C; 37pp; English.
 CC This represents the nucleotide sequence of an artificial transposon AT-2
 CC plasmid (part-2). The invention provides methods for creating artificial
 CC transposons and inserting these transposons into DNA targets in vitro.
 CC The artificial transposon consists of a linear blunt-ended DNA molecule
 CC comprising a marker gene flanked by terminal sequences, each comprising
 CC at least 4 bp of a U3 sequence of yeast retrotransposon Tyl. The DNA
 CC molecule used for generating such artificial transposons comprise an
 CC origin of replication; a first selectable marker gene; two blunt-ended
 CC transposon terminal, each of at least 4 bp, that are substrates for yeast
 CC retrotransposon Tyl integrase and flank a first restriction site useful
 CC for insertion of a second selectable marker gene to form an artificial
 CC transposon, and a second restriction site flanking the two transposon
 CC terminal, where digestion with the second restriction enzyme liberates a
 CC blunt-ended fragment having the transposon terminal at either end, the
 CC fragment being an artificial transposon. The method is useful for DNA
 CC mapping and sequencing of e.g. both functional and non-functional
 CC sequences, primer sites and restriction sites.
 XX
 SO Sequence 4933 BP: 1431 A; 1031 C; 1157 G; 1314 T; 0 other;
 Query Match 25.8%; Score 356; DB 20; Length 4933;
 Best Local Similarity 65.5%; Pred. No. 3.76-92;
 Matches 521; Conservative 0; Mismatches 275; Indels 0; Caps 0;
 QY 209 catgctcaggaatctttagcaggaagggccagagcaacatccgacgttctagaaa 268
 DB 416 catgctcaggaatctttagcaggaagggccagagcaacatccgacgttctagaaa 475
 QY 269 gctctgacatgacagagagaagaacacatctcgtcgtcccttgatgctggac 328
 DB 476 gctatttaatatatgacgaagaagaacaaacttgctgctcattggatctgac 535
 QY 329 gctagaagaactctgagagtagacagacacgctggagccgacatctgctgtagaac 388
 DB 536 caccagaagaaatctcgtgagtagatgaaagcatagatcccaaaattgcttaataaac 555
 QY 389 acatgctgacatactgacggaacttcgacatcgagacagacatcaagccgctgacagact 448
 DB 596 acatgtagatactctgactgattcttcacatggagggacagcttaagccgctaaagcatt 655
 QY 449 tgggagctaaagacacactcgtcttcctgagagccgacagctgctgacttgacacac 508
 DB 656 atccgcaagtaacaaatttctactctcgagacagaacaaattgctgacttgatgac 715
 QY 509 ggttaagctgacatctcctcgcgctgacatccgctatccgagtgaggatattaccaa 568
 DB 716 agtcaaaatgacgactcctgctgctgctgctgctatcagaaatagacgaatggcgacacttcgaa 775
 QY 569 tgcacacgctgacacgagcccgctgtagacgagctgaaagagctgacgaactggc 628
 DB 776 tgcacacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 835
 QY 629 ctcaagaggaacccagaggggtgctgctgctgctgctgctgctgctgctgctgctgctgct 688
 DB 836 aacaaaggaacccagagggcttctgctgctgctgctgctgctgctgctgctgctgctgct 895

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2001, 21:40:18 ; Search time 81.04 Seconds

(without alignments)
2744.339 Million cell updates/sec

Title: US-09-582-779A-1

Perfect score: 1380

Sequence: 1 ccgagcgaactcattggaag.....aataatgacgcttgaaattc 1380

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 280836 segs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/lna/5A_COMB.seq:*

2: /cgn2_6/prodata/1/lna/5B_COMB.seq:*

3: /cgn2_6/prodata/1/lna/6_COMB.seq:*

4: /cgn2_6/prodata/1/lna/PCTUS_COMB.seq:*

5: /cgn2_6/prodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	370.4	26.8	2330	2	US-08-557-128-3
2	357	25.9	5181	3	US-08-801-344-5
3	356	25.8	4164	1	US-08-204-675-1
4	356	25.8	4164	2	US-08-660-754-1
5	356	25.8	4164	2	US-08-996-364-1
6	356	25.8	4164	4	PCT-US95-02520-1
7	356	25.8	4933	1	US-08-204-675-2
8	356	25.8	4933	2	US-08-660-754-2
9	356	25.8	4933	2	US-08-796-364-2
10	356	25.8	4933	4	PCT-US95-02520-2
11	356	25.8	7102	3	US-09-138-024-20
12	356	25.8	7333	3	US-09-138-024-21
13	356	25.8	7633	3	US-09-028-851-1
14	356	25.8	7633	3	US-08-815-520-1
15	352.8	25.6	1115	1	US-08-507-455-5
16	349	25.3	1220	5	5204252-1
17	336.4	24.4	906	5	5204252-3
18	289.4	21.0	2688	1	US-08-088-633-3
19	289.4	21.0	2688	1	US-08-245-756-3
20	289.4	21.0	2688	1	US-08-441-750-3
21	289.4	21.0	2688	2	US-08-441-751-3
22	289.4	21.0	2688	4	PCT-US92-02521-3
23	115.2	8.3	8533	1	US-07-845-989-6
24	115.2	8.3	8533	1	US-07-845-989-6
25	39.4	2.9	7218	1	US-08-232-463-14
26	38.4	2.8	1931	2	US-09-130-114-2
27	38.2	2.8	2779	3	US-08-482-677-5
28	38.2	2.8	2783	1	US-08-152-019A-41

C 29	36.6	2.7	43280	2	US-08-804-227C-1	Sequence 1, Appl1
C 30	36	2.6	24417	2	US-08-846-762-1	Sequence 1, Appl1
C 31	35.8	2.6	20235	3	US-07-642-734C-3	Sequence 3, Appl1
C 32	35.8	2.6	20235	3	US-08-439-009A-3	Sequence 3, Appl1
C 33	35.4	2.6	1182	1	US-08-241-943-23	Sequence 23, Appl1
C 34	35.4	2.6	1431	1	US-08-254-357-1	Sequence 1, Appl1
C 35	35.4	2.6	2327	5	5229279-5	Patent No. 5229279
C 36	35.4	2.6	2327	5	5229279-5	Patent No. 5229279
C 37	35.4	2.6	4983	1	US-08-472-358-1	Sequence 1, Appl1
C 38	35.4	2.6	4983	4	PCT-US92-0576A-1	Sequence 1, Appl1
C 39	35.4	2.6	4984	1	US-08-687-806-1	Sequence 1, Appl1
C 40	35.2	2.6	3061	1	US-08-700-576-1	Sequence 1, Appl1
C 41	34.4	2.5	28958	1	US-08-258-261B-6	Sequence 6, Appl1
C 42	34.4	2.5	28958	1	US-08-456-837-6	Sequence 6, Appl1
C 43	34.4	2.5	28958	1	US-08-457-342-6	Sequence 6, Appl1
C 44	34.4	2.5	28958	1	US-08-457-342-6	Sequence 6, Appl1
C 45	34.4	2.5	28958	1	US-08-458-076A-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-08-557-128-3
Sequence 3, Application US/08557128
Patent No. 5849524
GENERAL INFORMATION:
APPLICANT: KONDO, Keiji
APPLICANT: KATIMARA, Susumu
TITLE OF INVENTION: TRANSFORMATION SYSTEMS FOR THE YEAST
TITLE OF INVENTION: CANDIDIA UTILIS AND THE EXPRESSION OF HETEROLOGOUS GENES
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,128
FILING DATE: 25-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP95/01005
FILING DATE: 25-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-129287
FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-285823
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-135015
FILING DATE: 25-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2330 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1259..2059
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1259..2059
; US-08-557-128-3

```

```

Query Match      26.8%; Score 370.4; DB 2; Length 2330;
Best Local Similarity 67.5%; Pred. No. 3.3e-101;
Matches 521; Conservative 0; Mismatches 251; Indels 0; Gaps 0;

```

```

Qy 214 caacgaatcttaccgaagaagcgcaaacgacacacatctgcagcttgcagaaagcttc 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1266 CCACGTTATCGTACACAGAGACGCGATCGCACACCTTCGCACTGCTCAAGCGTCTGT 1325

Qy 274 tggcatttgacgagagaagaaccacatctcgcgcttcccttgaatgctggagactta 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 TTTCCGCTTATGAGTCCAAAGAGACGACCTGTGTGCCAGTGTCTGATCTGTACACAG 1385

Qy 334 gaaagctcttgagcttagacagacgctgagcgacacatctgtcgtctgaagacata 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 AGGAGTGTCTCAAGCTCGTGTATACGCTTGCTTATATCTGTCTGTGAAGACGCTTA 1445

Qy 394 tgcacatactgacgagacttcgacatcgagacgacagctcaagcgccttgacagacg 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1446 TTGATATATCTGATGACCTCTCTATGAGATCTACTGTGCTCACTGTGAGCTTTCAA 1505

Qy 454 ctaagacacactatcatgactcttcgagagacgcaagctcgcgacattgcaacagctta 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1506 ACAAGACAAATTTCTCATCTTTGAGGACCGTAAAGTTTGATTCGCAACACCGTCA 1565

Qy 514 agctcgagtaacctccctccgctgctacacgctatcgagtgagtgagcagatactaac 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1566 AGGCACATGACCGCGGTGTGCTTCAAGATTGCCCATGGCAGATATACACCAACGCC 1625

Qy 574 acgagctacacgccccgctgtatagacgagctgaagaagcttgacgaactgagctaac 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1626 ACGGTGTACACGGTGCAGGTATCGTCAAGGGGTGAAGAGGCTGCACAGAAACACAG 1685

Qy 634 aggaacacgaggggtgtcgtatgctgagcagagctctctcgaagggctcttgagcgcg 693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1686 ATGAGCCAAAGAGGGCTGTGTGCTTGCGGAGCTGAGCTCCAAAGGGCTCTTGCGCC 1745

Qy 694 gaaacataccgagcgctgctgaatgagcgaagcttgagacgaagacttctgacaggt 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1746 GGCATATATACCGAGGAGACCGTGGAGATTGCCAAAACTGATAGAGACTTTGTATTG 1805

Qy 754 tcatcgagcagcgctgacatgggtgagcgctgacagagcttgaactgagctcatcatgccc 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1806 TCATCCACAGAGAGACATGCGTGCAGAGAAATGGGTTGCGATGATCATCATGAGAC 1865

Qy 814 cgggggttgcttgacgacgaagaagagcgcttgagcgaagctacgacgaagcgatg 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1866 CAGGCTTGAGACTTCGATGAAGGGGAGCTCCCTGGGCCAACAAGTATGTCGATG 1925

Qy 874 aggtctgacgagcgctacacgctgcatcatcttgacgaagggctcttgacgaagaa 933
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1926 AGGTTTCACTGCTGTGCTTGCATCATCATCTGTGTGATGAGGCTTTGTTGAAAGG 1985

Qy 934 gaaacacacgaagtgacagggctgacgctacacgaagcgcttgaggaagctta 985
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1986 GAGATCCAAAGTGAAGAGTGAAGCGTTATAGAAAAGCAGGCTGGGATGCTTA 2037

```

```

RESULT 2
US-08-801-344-5/c
; Sequence 5, Application US/08801344
; Patent No. 6087140

```

```

; GENERAL INFORMATION:
; APPLICANT: Cameron, Douglas C.
; APPLICANT: Shaw, Anita J.
; APPLICANT: Alaias, Nedim E.
; TITLE OF INVENTION: MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL
; TITLE OF INVENTION: FROM SUGAR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dewitt Ross & Stevens S.C.
; STREET: 8000 Excelsior Drive, Suite 401
; CITY: Madison
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53717-1914
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,344
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sara, Charles S.
; REFERENCE/DOCKET NUMBER: 09820.037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-831-2100
; TELEFAX: 608-831-2106
; INFORMATION FOR SEO ID NO: 5
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; MOLECULE TYPE: circular
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Yeast shuttle vector YEP352
; US-08-801-344-5

Query Match      25.9%; Score 357; DB 3; Length 5181;
Best Local Similarity 62.7%; Pred. No. 5.5e-97;
Matches 555; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

```

```

Qy 120 acatcatcttgatgtgtgtaggtccatccacagtaagcattgtataagctgatacc 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1257 ATATACCATATGTATGTATGTAAGAAACATGAATTCGCCAGTATCTTAACCAACTGC 1198

Qy 180 atagggtctacgacacttagcacttgccacatctcaagaactcttagcagaagaagggc 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1197 ACAAGACAAAACCGAAGACGATTAATCATGTGCAAGCTCATATTAAGAACTGCT 1138

Qy 240 aagcagacacacttcgacagcttgtagaagctctcgtgcatgtagacgagaaagaaac 299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1137 GCTACTCATCTTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1078

Qy 300 aatctcgtgctccctccttgatgtagcgagcgtcagaagctcttgagcttagcagaag 359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1077 AACTTGTGTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1018

Qy 360 ctggagcgcacacttgctcgtgtagaagacatgtagacactagcagcgtcagacatc 419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1017 TTAGTCCCAAAATTTGTTACTTAAACACATGTGATGATGATGATGATGATGATG 958

Qy 420 gagaacgaagtcagcgcgtgacgacgctgagcgtcgaagcagaactcaatgactctcag 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 957 GAGGACACAGTTAAGCCGCTTAAGGCAATTATCCGCAAGTACATTTTACTCTTCGAA 898

Qy 480 gaccgcaagctcgtgacatgtagcagacagcttaagctcagtaactcctcggcggtgac 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Db 1196 GAGAGATCGCGCAG 1211

RESULT 8

US-08-660-754-2

Sequence 2, Application US/08660754

Patent No. 5843772

GENERAL INFORMATION:

APPLICANT: Devine, Scott E.

APPLICANT: Boeke, Jef D.

APPLICANT: Brateman, Lellita T.

TITLE OF INVENTION: In Vitro Transposition of Artificial

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie, and Beckett

STREET: 1001 G Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION NUMBER: US/08/660,754

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/204,675

FILING DATE: 02-MAR-1994

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 01107.45501

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202.508.9100

TELEFAX: 202.508.9299

TELEX: 197430 BMB UT

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 4933 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOHETICAL: NO

ANTI-SENSE: NO

IMMEDIATE SOURCE:

CLONE: PAT-2

US-08-660-754-2

Query Match 25.8%; Score 356; DB 2; Length 4933;
Best Local Similarity 65.5%; Pred. No. 1.1e-96;
Matches 521; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

QY 209 catgtcaagaatctttagcagagaagagccacgaatccagcttgctgaagaa 268
DB 416 CATGTCAAGAAAGCTATATTAAGAGAGCTGCTGCTACTCATCTAGTCTGCTCCAA 475
QY 269 gctctcgtcagatgatgacagagaagaacaaatctctcgtcctccctgagtgagac 328
DB 476 GCTATTTAATATCATGACGAGAAAGCAAACTGTGTGCTTCACTTGTGATGTTGCTAC 535
QY 329 gctcagaagctctgagagctgacagacagctggagaccgacatttgcctcgaagac 388
DB 536 CACCAAGGAAATTAATCTGAGCTAGTGAACCAATTAAGTCCCAAAATTTGTTACTAATAAAC 595
QY 389 acatgtcgcatactcgaagacttcgacatcgagacgacagcagcgctcgaagact 448

Db 596 ACATGTGGAATATCTTACATGATTTTTCATGTGAGGACACAGTTAAAGCCATTAAGCATTT 655

QY 449 tgcgctcgaagacaaatcatalgatctcgaagacagcgaagtctgcgtacatctgcaaac 508

Db 656 ATCCGCCCAAGTACAAATTTTACTCTTCGAGACAGAAATTTGCGACATTTGTAATAC 715

QY 509 ggttaagctgagactcctcctcggcggtgacagctatcggaggtgggagatataccaa 568

Db 716 AGTCAATTTGCTACTCTGCGGCTGTATACGAATAGCAATGCGAGACATTAACAA 775

QY 569 tgcacagcgctcaccgcccggctgtaacagcggtcgaagagagctgcgaactgac 628

Db 776 TGCACACGGGTGGTGCGGCCAGATTTCTTAGCGTTTGAAGACGCGCGCAAGAAAGT 835

QY 629 ctacagagaacccaggggtgctgacgtcgaagagctctctcgaaggtcttgac 688

Db 836 AACAAAGAACCTAGAGGCTTTGTATGACCAATTTGATCCAAAGGCTCCCTATC 895

QY 689 gcgcgagagactaacgcggtgctgtaaaatgacagagcgcgagagacttgat 748

Db 896 TACTGAGATATATCTAAGGCTACTGTTGACATTTGAGAGACGACAAAGATTTGTTAT 955

QY 749 cgggttcacgcgcagcgtgacatggtggtgctgcaagagcgttgactgctcatcat 808

Db 956 CGGCTTTATTTGCTCAAAAGACATGGGTGGAAGATGAAAGTTGATTTGATTTAT 1015

QY 809 gacccgggggttgcctcctgacagacaagagagagcctgggacagcagctacagcgt 868

Db 1016 GACACCGGTGTGGTTTATGATGACAGAGGACGACCATTTGCTCAACGTTATGAAACGT 1075

QY 869 ggtatgagtcgtcagcgaagctacatgtaacatctgtgcaagagcctcttgacaa 928

Db 1076 GGATGATGTGTCCTACAGATCTGACATTTATTTGTTGGAAGACATTTTGCAAA 1135

QY 929 ggtgaagaagaccccaaggtctgaggtgcccgtacacgaagcggcttgaggacttact 988

Db 1136 GGAAGGAGTGTCTAAGGTGAGAGGTGAACGTTACAGAAAGACGCTGGAAACATATTT 1195

QY 989 gcgcgctatggcgag 1004

Db 1196 GAGAGATCGCGCAG 1211

RESULT 9

US-08-796-364-2

Sequence 2, Application US/08796364

Patent No. 5968785

GENERAL INFORMATION:

APPLICANT: Devine, Scott E.

APPLICANT: Boeke, Jef D.

APPLICANT: Brateman, Lellita T.

TITLE OF INVENTION: In Vitro Transposition of Artificial

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie, and Beckett

STREET: 1001 G Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/796,364

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/204,675


```

FILE REFERENCE: 0342/1D469US1
CURRENT APPLICATION NUMBER: US/09/138.024A
CURRENT FILING DATE: 1998-08-21
EARLIER APPLICATION NUMBER: 60/056,719
EARLIER FILING DATE: 1997-08-22
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 21
LENGTH: 7333
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: plasmid pZM197
US-09-138-024-21

```

```

Query Match      25.8%; Score 356; DB 3; Length 7333;
Best Local Similarity 65.5%; Pred. No. 1.3e-96;
Matches 521; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

```

```

OY 209 catgtcaacgaatctctacgacgaagggcgaagcacaacatctgcagttgtctagaac 268
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
DB 6045 catgtcgaaggtacataataaggaacgctgcgtactctactctactgtctgtccaa 6104
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 269 gctctgtcgtatgtatgcacgaggaagaaacaaatctctgcgtctccctgtatgtcgagc 328
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6105 gctacttaatacagacgacgaagaaacaaactgtgtctctcatgtatgtatgtctgtac 6164
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 329 gctctgaagagctctctgtgagctgcagacacgctgtgagccacatctgtctgcgaagac 388
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6165 caccagaagaattactcgtgagctgtctgtgagcattgtcccaaatctgtctactcaaac 6224
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 389 acatgtcgaacatactacgagacttcgacacacacacacacacacacacacacacacacac 448
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6225 acatgtgatactctgacgacttctcactgagcagacacacacacacacacacacacac 6284
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 449 tgcgtctgaagcacaactcatatgactctctgagagacacacacacacacacacacacac 508
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6285 atccgcgaagtaacatttctctctcgaagacgaagaaatctgtctgacattgtgaaac 6344
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 509 ggttaagctcgaactctctcctcctcctcctcctcctcctcctcctcctcctcctcctc 568
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6345 atctcaatctgacgtactctcctcgtgctacacgaatgacgaatgtgagacacattacga 6404
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 569 tgcacacggtcgtacacgacccgctgtgtacacacacacacacacacacacacacacac 628
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6405 tgcacacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6464
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 629 ctcaacggaacacacacacacacacacacacacacacacacacacacacacacacacacac 688
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6465 aacaaaggaacacacacacacacacacacacacacacacacacacacacacacacacacac 6524
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 689 ggcggaagacataacacgacggtcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 748
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6525 tacttgagagataatacgaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6584
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 749 cgggttcatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 808
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6585 cgggttcatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 6644
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 809 gaccccgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 868
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6645 gacacacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6704
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 869 ggttaagctcgaactctctacgacacacacacacacacacacacacacacacacacacac 928
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6705 ggttaagctcgaactctctacgacacacacacacacacacacacacacacacacacacac 6764
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 929 ggggaagagacacacacacacacacacacacacacacacacacacacacacacacacacac 988
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 6765 ggggaagagacacacacacacacacacacacacacacacacacacacacacacacacacac 6824
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 989 ggcgcgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1004
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||

```

DB 6825 gagaagatgcggccag 6840

```

RESULT 13
US-09-028-851-1/C
Sequence 1, Application US/09028851
Patent No. 6060249

```

GENERAL INFORMATION:

```

APPLICANT: Kevin Baker, Austin Gurney
TITLE OF INVENTION: Method of selection for Genes Encoding
TITLE OF INVENTION: Secreted and Transmembrane Proteins
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 1 DNA Way
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

```

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 Inch, 1.44 Mb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/028,851

FILING DATE: 26-Feb-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/039476

FILING DATE: 27-Feb-1997

ATTORNEY/AGENT INFORMATION:

NAME: Svoboda, Craig G.

REGISTRATION NUMBER: 39,044

REFERENCE/DOCKET NUMBER: P1058R1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650/225-1489

TELEFAX: 650/952-9881

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7633 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

US-09-028-851-1

Query Match 25.8%; Score 356; DB 3; Length 7633;

Best Local Similarity 65.5%; Pred. No. 1.4e-96;

Matches 521; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

```

OY 209 catgtcaacgaatctctacgacgaagggcgaagcacaacatctgcagttgtctagaac 268
    |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||  |||||  ||
DB 3768 catgtcgaaggtacataataaggaacgctgcgtactctactctactgtctgtccaa 3709
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 269 gctctgtcgtatgtatgcacgaggaagaaacaaatctctgcgtctccctgtatgtcgagc 328
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 3708 gctatgttaataatcgtacacacacacacacacacacacacacacacacacacacacacac 3649
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 329 gctctgaagacacacacacacacacacacacacacacacacacacacacacacacacacac 388
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 3648 caccagaagaattactcgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3589
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 389 acatgtcgaacatactacgagacttcgacacacacacacacacacacacacacacacacacac 448
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 3588 ACATGTGATATCTTGAATATTTTTCATGAGGCGACAGTTAAAGCCGTAAAGCATTT 3529
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 449 tgcggtcaagcacaactctatgactctcgaagacacacacacacacacacacacacacacac 508
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
DB 3528 ATCCGCAAGTACATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3469
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
OY 509 gtttaagctcgaactctctcctcctcctcctcctcctcctcctcctcctcctcctcctc 568
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||

```

Db 3468 AGCTAAATTCAGTACTCTCGGGTGTATACGAATAGCAGAAATGGGACAGACATTACGAA 3409.
Qy 569 Tgcacagcgctcaccgcccgtgtgatagccggcctgaagagcctcgaacacg 628
Db 3408 TGCACACGGCTGTGTGGGCCAGGATTGTTAGCGCTTTTAAAGAGCGCGCAGAGAAAGT 3349
Qy 629 ctcaagaaacccaggggtctgtatctgtcagagctctctctcagggcctcttg 688
Db 3348 AACAAAGAACTAGAGCGCTTTTGTATGTATAGCAATTCATGCAAGGCGCTCCATC 3289
Qy 689 ggcggagactatacccgccggtcgtgtgaatggcgagccttgagaaactctgtat 748
Db 3288 TACTGAGAAATATCTAAGGCTACTGTTGACATTTGCGAAGAGCGCAAAATTTGTTAT 3229
Qy 749 cgggttcacgcagcgctgtacatgggtgggtggcctgacagagccttgacgtctcat 808
Db 3228 CGGCTTATTTGCTCAAGAGACATGGGTGCAAGAGATGAAAGTTACCATTTGTTAT 3169
Qy 809 gaacccgggggttgccctgacagaaagagacgctggcctgacagcagctacgcag 868
Db 3168 GACACCCGGTGTGGGTTAGATGACAAGGAGAGCGATTTGGTCAACAGTATAGAACCT 3109
Qy 869 ggttgagctgtcagcagcgtatccgtatgcatctgttgcaagagcctcttgaca 928
Db 3108 GGATGATGTGCTCTACAGGATCTGACATTTATTTGGAAGAGGACTATTTGCAAA 3049
Qy 929 gggagagaccccaaggctgaggggtgcccgtacccagagcgggttgagagcctact 988
Db 3048 GGGAGAGGATGCTAAGGTAGAGGCTGACACTTACAGAAAGCAGGCTGGAGACATATTT 2989
Qy 989 ggcgcgtatggcgag 1004
Db 2988 GAGAGATGCGGCCAG 2973

RESULT 14
US-08-815-520-1/c
: Sequence 1, Application US/0815520A
: Patent No. 6136569
: GENERAL INFORMATION:
: APPLICANT: Kevin Baker, Austin Gurney
: TITLE OF INVENTION: Method of Selection for Genes Encoding Secreted and
: TITLE OF INVENTION: Transmembrane Proteins
: FILE REFERENCE: P1058
: CURRENT APPLICATION NUMBER: US/08/815,520A
: CURRENT FILING DATE: 1997-02-27
: NUMBER OF SEQ ID NOS: 3
: SEQ ID NO 1
: LENGTH: 7633
: TYPE: DNA
: ORGANISM: Artificial
: FEATURE:
: NAME/KEY: Artificial Sequence
: LOCATION: 1-7633
: OTHER INFORMATION: Amylase yeast expression plasmid
US-08-815-520-1

Query Match 25.8%; Score 356; DB 3; Length 7633;
Best Local Similarity 65.5%; Pred. No. 1.4e-96;
Matches 521; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

Qy 209 catgtcaagaaatcttaccgagaaaggccaaaggacacaaatctgcagctgttagaaa 268
Db 3768 CATGTCAAGAACTATATAGAGAGCTGCTACTACTCTAGCTCTGTTGCTGCCAA 3709
Qy 269 gcttcgcatctgacagagaaacaaatctcctgcctccctcgatgagctgcagc 328
Db 3708 GCATTTATATCATCATCAGCAAGAAAGCAAACTTGTGTCTCTCTTGTGATGCTGTC 3649
Qy 329 gcttagaagcctctgagctagcagacagcgtggagccgacatctgtcctgaagac 388

Db 3648 CACCAAGAAATTACTGAGTATGTTGAACATTAGTCCCAAAATTTGTTACTAAAC 3589
Qy 389 acatgtcagacatactgagcacttcgacacacagagacagcctgcagcagct 448
Db 3588 ACATGTGATATCTTGTACATGATTTTTCATGAGGAGCAGATTAAGCCGCTAAAGCAAT 3529
Qy 449 tgcgctlaagcacaactcatalctctcgaagccgaagctcgtcgtcatctgcaaac 508
Db 3528 ATCCGCAAGTACAAATTTTACTCTCTCAAGACAGAAATTTGCTGACATTTGTAATAC 3469
Qy 509 ggtlaagctgacactcctccgagctgtacccgtacgcagagtgagcagatataccaa 568
Db 3468 AGTCAAAATTTGCACTCTCGCGGTGTATACAGAAATAGCAGAAATGGCAGACATACCA 3409
Qy 569 tgcacagcgctcaccgcccgtgtgtatagccggctggaagagccttgcaaacctgc 628
Db 3408 TGCACACGGTGTGTGGGCCAGGATTTGTTAGCGGTTTGAAGCAGGCGCAGAGAAAGT 3349
Qy 629 ctcaagaaacccaggggttgctgtatctgtgcaagcctctctcagggcctctggc 688
Db 3348 AACAAAGAACTAGAGCGCTTTTGTATGTATAGCAATTTGCTATCAAGGCGCTCCATC 3289
Qy 689 ggcggagactatacccgccggtcgtgtgaatggcgagccttgagaaactctgtat 748
Db 3288 TACTGAGAAATATCTAAGGCTACTGTTGACATTTGCGAAGAGCGCAAAATTTGTTAT 3229
Qy 749 cgggttcacgcagcgctgtacatgggtgggtggcctgacagagccttgacgtctcat 808
Db 3228 CGGCTTATTTGCTCAAGAGACATGGGTGCAAGAGATGAAAGTTACCATTTGTTAT 3169
Qy 809 gaacccgggggttgccctgacagaaagagacgctggcctgacagcagctacgcag 868
Db 3168 GACACCCGGTGTGGGTTAGATGACAAGGAGAGCGATTTGGTCAACAGTATAGAACCT 3109
Qy 869 ggttgagctgtcagcagcgtatccgtatgcatctgttgcaagagcctcttgaca 928
Db 3108 GGATGATGTGCTCTACAGGATCTGACATTTATTTGGAAGAGGACTATTTGCAAA 3049
Qy 929 gggagagaccccaaggctgaggggtgcccgtacccagagcgggttgagagcctact 988
Db 3048 GGGAGAGGATGCTAAGGTAGAGGCTGACACTTACAGAAAGCAGGCTGGAGACATATTT 2989
Qy 989 ggcgcgtatggcgag 1004
Db 2988 GAGAGATGCGGCCAG 2973

RESULT 15
US-08-507-455-5
: Sequence 5, Application US/08507455
: Patent No. 5695961
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: BIFUNCTIONAL EXPRESSION SYSTEM
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NIXON & VANDERHAYE P.C.
: STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
: CITY: ARLINGTON
: STATE: VIRGINIA
: COUNTRY: USA
: ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/507,455
FILING DATE: 08-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9303988.1

FILED DATE: 26-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: MITCHARD, LEONARD C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 1498-72
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
US-08-507-455-5

Query Match 25.6%; Score 352.8; DB 1; Length 1115;
Best Local Similarity 65.2%; Pred. No. 4.1e-96;
Matches 519; Conservative 0; Mismatches 277; Indels 0; Gaps 0;

```
OY 209 catgtcaacgaatcttcaacgagaagggccaaagcacacaacttcgcaattgctagaac 268
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 233 CATGTGGAAGACTACTATATAGGAAGTGTCTACTACTCTCTAGTCTCTGCTGCCAA 292

OY 269 gctctcgtacatgtagacgagagaagaacaaactctcgtcccttgatgtgcgagc 328
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 293 GCTATTATTATCATCATGCAAGAAACAAACTTGTCTCTCATGTGATGTTCTGTAC 352

OY 329 gctcgaagaagctctcgtgaagcttagacacgcgtcgtgacccgacatttgtcgtgaagc 388
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 353 CACCAAGGATTACTGAGTGTAGTGAAGCATTTAGTCCCAAAATTTGTTACTAAAGC 412

OY 389 acaatgtcacatacttcaacgagcttcgacatcgagacgacagtcgaagccgtcgaagcagct 448
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 413 ACATGTGATATCTTGATCTGATTTTTCATGAGAGGACACGTTAAGCCGCTAAGGCAAT 472

OY 449 tgcggtcgaagcaaacatcactgactcgtgaggaacgcgcaagtlcgctgacattgccaacg 508
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 473 ATCCGCCAAGTACAAATTTTCTACTCTTCGAGAGACAGAAATTTGCTGACATTGTGTAATAC 532

OY 509 ggttaagctcgagctactcctccggtcgtagccgtatcgcgaggtggtgcgatatcaaca 568
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 533 AGTCAAAATTGCACTCTGCGGGTGTCTATAGAAATGCAAGATGGCGACATTAACGAA 592

OY 569 tgcacacgcgtcaccggtcccggtgtgaaagcgtgaaggaagcgtgcgaagactgac 628
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 593 TGCACACGGGTGTGTGGGCCCAAGATTTGTAGCGGTTTGAAGCAGCGCGCAAGAAAGCT 652

OY 629 ctacacagaaccccaagggtgtcgtgaagtcgtgcaagagctcctctcgaaggtccttgagc 688
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 653 AACAAAGGAACCTAGAGACTTTTGTATGTAGCAGAAATGTCATGCAAGGCTCCCTATAC 712

OY 689 ggcggtgagactataaccggtggtcggtgtaaatgtggaagcgtggaagagacttgtgat 748
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 713 TACTGGAGATATACTTAAGGCTACTGTGACATTCGCAAGACGCAAGAAATTTGTAT 772

OY 749 cgggttcaatcgcaagcgtgacatggtgtggtggtggtggtggtggtggtggtggtggtggt 808
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 773 CGGCTTTATTGCTCAAGAGACATGGGTGAGAGATGAAAGTTAGATGCTGTTGATTAT 832

OY 809 gaacccggtggtgtggtcgtggaacgaagaagagcgtggtggtggtggtggtggtggtggtggt 868
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 833 GACACCCGGGTGTGTGTAGATGCAAGGAGACGCTTGGGTCAACAGTATAGAACCT 892

OY 869 ggttgaggtgtcgtgagcaggtacagatgtgtaacatgtgtggtggtggtggtggtggtggtggt 928
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 893 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 952

OY 929 ggtgaagaagaccccaagggtcgtgaggtgtccgtctacgcgaagcgtgtggtggtggttactt 988
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
DB 953 GCGAAGGATGCTAAGCTAGAGCGGTGACGTTACAGAAAAAGCAGCTGGGAAGCATATTT 1012
OY 989 ggcgcgtatgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 1004
DB 1013 GAGAAGATGGGCGCAG 1028
```

Search completed: April 4, 2001, 00:02:04
Job time: 8506 sec

Wed Apr 4 11:25:31 2001

us-09-582-779a-1.rni

Page 13

Oy	291	aagaaacaaatctctcgtgcctcttgatgtagcgctctagaagcctctgagagcta	350
Db	10	AAACAGAGAACCTCTGCTTGGCCCGACCTGCACACCCCGATGAGATCTGGAGCTG	69
Oy	351	gcagacacgcctgaggaccacatctgtcctctgaagacacatctgcacatactgacgac	410
Db	70	GCCGACAGTGGCGGTCCCTTACATTTTGGCTGTGAAAGACGACGTAGACATTTGGAGAT	129
Oy	411	ctcgacalcgagacgacagctcaagccgctcgagcagctctgcygctaaagcaactcaty	470
Db	130	TT---CACTGATTAATTCATCATCGTAGACCTGCACACTCTGGCCGACGACCAATTTCTG	186
Oy	471	atctctcgagagaccggaagcttcgtctgcaaatctggaacacggtctaaagctgaactctcc	530
Db	187	CTGATGAGGAGATCCGCAGATTTCACAGACTCGGCACACGGTGTCTCCCTGCAGTACGGCAG	246
Oy	531	ggcgcgtatccglatccgagtagtgaggcgataataccaatgacaacagcgctacacgagccc	590
Db	247	GGCATCTTAAAGATTTCAGTTGGGGCAGATCTGTCTACAGGCCACACACTTACCTGGACGT	306
Oy	591	ggctgacagccggagctgaagagga	615
Db	307	AGTATTCTGCAGGGCTCTGAAGCGG	331

RESULT	11
BE203967	
LOCUS	610 bp mRNA
DEFINITION	EST39666.3 KVO Medicago truncatula cDNA clone pKVO-13J7, mRNA sequence.
ACCESSION	BE203967
VERSION	BE203967.1 GI:8747248
KEYWORDS	EST.
SOURCE	Daniel medic.
ORGANISM	Medicago sativa

REFERENCE
AUTHORS

Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta: Magnoliopsida: eudicotyledons: core eudicots: Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Medicago.

1 (bases 1 to 610)

Vandenbosch, K., Endre, G., Hur, J., Moore, J., Beremand, P., Ellis, L., Town, C. D., Bowman, C. L., Craven, M. B., Hansen, T. S., Holt, I. E. and Fraser, C. M.

TITLE	ESTs from uninoculated seedling roots of <i>Medicago truncatula</i>
JOURNAL	Unpublished (1999)
COMMENT	Contact: VandenBosch K

Email: katemall@bio.tamu.edu
 Texas A&M University: 76763739e
 TIGR sequence name: MTGAK522K
 More information is available at:
<http://chrysis.tamu.edu/medicago>
 Seq primer: Skmod (CTA gaa CTA gtc gat CC)
 Location/Qualifiers

```

1. 610
source
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pKV0-13j7"
/clone_lib="KV0"
/tissue_type="Seedling roots"
/dev_stage="Immediately prior to inoculation with
Sinorhizobium meliloti (0 hour)"
/lab_host="E.coli strain XL04r"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist

```

helper phage and propagated in XL0LR cells.				
BASE COUNT	181 a	92 c	164 g	173 t
ORIGIN				

Query Match	7.18;	Score 97.4;	DB 105;	Length 610;
Best Local Similarity	55.88;	Pred. No. 2.2e-17;		
Matches 207; Conservative	0;	Mismatches 161;	Indels 3;	Gaps 1

QY	222	tcctacgcagaagaagggcccaagacacacatctgcacgctctcgaagaagctcttcggcattg	28
Db	237	TCATTTGGGGAGAGCGCGCAATCTGTCMCACAGAAATCAACAGAGAAAGAGTTGTTTGAGATA	296
QY	282	atgcacgcagaagaacaaccaatctctgcgctctcccttattgtgcygaagcttagaagctc	341
Db	297	ATGGCTCGACGAGGAGACAGTAATCTGTGTGTTTGCGCTGCACATGTGTGAACTCGAAGCTGAATTG	356
QY	342	ctggagctcgaagcagacacgctctggagccgcacattctctgcctgaagacacactctgcacata	401
Db	357	CTTGAATAATGCCGGAAGGTTGGACCTCGACATATGCTTGTCTGGAAGACTCATGTGGATATT	416
QY	402	ctgaagcagcttcgacatcgaagacagacagltcaagccgctctgcacagctctgcgcttgaagcac	461
Db	417	TTGGCTGATTTTACTCGCGA---TTTGGCTCTPAAGCTTCTCTCGATTTGCAAGAAACAT	473
QY	462	aactctcatgatctcttcgaggaacgcgaagltcgcctgcacattggcaacacggtltaagctgcag	521
Db	474	AACCTCTTAATCTTTGAGAGATCGTAAATTTGCTGACATTTGGTAAACGTAACCATTCGAA	533
QY	522	tactctctccgcgctgtaccgctatctcgcggaagcggagatattaccaatggacaacgcgsgctc	581
Db	534	TATGCAAGAGGAGATCTTTCACATATTGGATTGGCGCTCACATAGTAAATGTCTCACATTAATC	593
QY	582	accggaacccgcg	592
Db	594	TCTGGTCTCTGG	604

RESULT	12
AL369478	
LOCUS	AL369478 461 bp mRNA
DEFINITION	MEAI31D04.F1 MBIA Medicago truncatula cDNA clone MCBAI31D04 T3, mRNA sequence.
ACCESSION	AL369478
VERSION	AL369478.1 GI:9669231
KEYWORDS	EST.
SOURCE	barrel medic.
ORGANISM	Medicago truncatula

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 461)	Journet, E. P., Crespeau, H., van Tuinen, D., Gouzy, J., Jallion, O., Nibbel, A., Carreau, V., Chatagnier, O., Kahn, D., Glanina, Zl-Pearson, V. and Gams, P.	Medicago truncatula ESTs from nitrogen-starved roots	Unpublished (2000)	Contact: Genoscope

Genoscope - Centre National de Séquençage
BP 191 91006 Evry cedex - France
Email: segrete@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Moléculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
Mc-est@toulouse.inra.fr Website :
<http://sequence.toulouse.inra.fr/Mcunucatula.html>).

Location/Qualifiers
I. .461

FEATURES
SOURCE

```

/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone="MCBA31D04"
/clone_lib="MTBA"

```


nystatin and cefotaxime in covered crystallization dishes. Roots were harvested. The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made in the IVI Close Lab (Choi, Close, Fenton) at the University of California, Riverside. The cDNA clones were in vivo excised to give pluscript phagemids before normalization and carried out. The mass excision of phagemid library and normalization were done in HT Nguyen lab by D. Zhang at Texas Tech University. Normalization protocol used was that of Soares. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors).

BASE COUNT 144 a 140 c 150 g 147 t
ORIGIN

Query Match 9.3%; Score 128.4; DB 108; Length 581;
Best Local Similarity 58.2%; Pred. No. 2e-26;
Matches 269; Conservative 0; Mismatches 181; Indels 12; Gaps 2;

Oy 300 aatctctgcttcccttgatgtgctgagcgtctagaagctcttgagctgacagacg 359
1 AACCTCTGGTGGCCGCGCCGATGTACAAACAGCAAGAACTCTTGACCTACCAACAG 60
Oy 360 ctggagccgacatctgtctctgaagacacatgctgacatactgagcgaactcgacac 419
61 GTTGGTCCAGATTTGCTGCTGAACCCATGTGATTTCTGTGATTTTACACCA 120
Oy 420 gagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 479
121 GATTTTGGTGCCAA--GCTCCGTTGATTTGCTGAAGACACACTTTTGTATCCTTGA 177
Oy 480 gaccgaagctctgagcagcagcagcagcagcagcagcagcagcagcagcagcag 539
178 GACCCGAATTTGCTGCTGACATGGGGAACGCTAATGCAATATGAAGAGAAATATTC 237
Oy 540 cgtatcgagagctgagcagcagcagcagcagcagcagcagcagcagcagcagcag 599
238 CCGATATACAGATGGCCGATATTTGACAGCGCATGTAGTCCCTGAGACTGGAATCA 297
Oy 600 gccggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 659
298 GATGCGTTGAACCTG-----AAGGGTTGCCAAAGGAAAGCGCTTACTTGGTT 348
Oy 660 gcaagcctctctctcagcagcagcagcagcagcagcagcagcagcagcagcagcag 719
349 GCTGAGATAGCGCCACGACGCAACCTTCTCATGGAGATTACACACGCGACCGTGAAG 408
Oy 720 atggcgaagcctggaagcagcagcagcagcagcagcagcagcagcagcagcagcag 761
409 TTGCGCGACCAACATCTGATTTGTTATCGGATTCCTTGTG 450

RESULT 5
BE394063 775 bp mRNA EST 21-JUN-2000
LOCUS 601312432F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3659383 5',
DEFINITION mRNA sequence.
ACCESSION BE394063
VERSION BE394063.1 GI:9339428
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 775)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LUCM349 row: p column: 08
High quality sequence stop: 652.
Location/Qualifiers

FEATURES

source

1..775
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3659383"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site:1: Choi; Site:2: EcORI; cDNA made by oligo-dT priming. Directionally cloned into EcORI/XhoI sites using the following 5' adaptor: GGCACGAG(5). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 197 a 157 c 224 g 197 t
ORIGIN

Query Match 8.4%; Score 116; DB 107; Length 775;
Best Local Similarity 56.7%; Pred. No. 8.9e-23;
Matches 279; Conservative 0; Mismatches 200; Indels 13; Gaps 3;

Oy 255 ccaatgctagaagcctctgacatgacagcagagaagaacacatctctgctcc 314
1 CCATTTGATGAGAGCTTCTCAGCGTTATGCAAAAGAGACCAACTGTGTATCT 60
Oy 315 ctgagtgtgagcgtctagaagcctctgagctgacagacgcctgagcagcagc 374
61 GCTGATGTTTCACGTGGCAGAGAGCTG-TGCAGCTACAGATGCTTTAGAGCTTATTC 119
Oy 375 tctgtctgaagacacatgctgacatgacagcagcagcagcagcagcagcagcagc 434
120 TGCATGCTGAAGACTCATGTGATTTTGAATGATTTTACTGCTGATGCA---TGAAG 176
Oy 435 ccgctgacagcagcctgagcagcagcagcagcagcagcagcagcagcagcagcagc 494
177 GAGTTGATTAACCTTGGCAAAATGCCATGAGTCTTGTATTTGAAGCCGGAATTTGCA 236
Oy 495 gaacttgcaacacagcgttaagcgtgacatctcccgcgctgtaacgctacgagagtg 554
237 GATATGGAACACAGTGAAGAAACAGCATATGAGAGAGATCTTTAAATAGCTTCTGG 296
Oy 555 gcgataatlaacaaatgacacagcagcagcagcagcagcagcagcagcagcagcagcagc 614
297 GCAGATCTAGTAATAGCTGACAGCTGCTGCAAGGCTCAGAGAGTGTGAAGCGCTGCAAG-- 354
Oy 615 gctgcgaacatgacctcaagaaacccaggggtgctgtagctcgtgagagcctctctc 674
355 -----AAGTGGCGCTGCTTTCATCGGGGGGCTCTCTTATTTGGAAATGAGCTCC 407
Oy 675 cagggtctcttgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 734
408 ACCGCTCCTCGCCACTGCGGACTACACTAGAGACGCGGTAGAAATGCGCTGAGAGACAC 467
Oy 735 gaagactctgt 746
468 TCTGATTTTGTG 479

RESULT 6
BE019969 557 bp mRNA EST 06-JUN-2000
LOCUS BE019969
DEFINITION sm38b04.y1 Gm-c1028 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1028-6152 5' similar to SM:PYR5_TOBAC Q42942 URIDINE

[illegible]

US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: andersen@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Strategene 5K primer.
Location/Qualifiers
1..579

/organism="Triticum aestivum"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="WHE1213_C11_E21"
 /clone_lib="Wheat etiolated seedling root cDNA library"
 /clone_type="Root"
 /dev_stage="Five day old etiolated seedling"
 /lab_host="E. coli SOLR"
 /note="Vector: lambda uni-ZAP XR, excised phagemid:
 Site_1: EcoRI; Site_2: XhoI. Seeds were surface-sterilized,
 germinated and grown aseptically in the dark at room
 temperature on filter paper with water, nystatin and
 ceftaxime in covered crystallization dishes. Roots were
 harvested. The tissue, total RNA, and poly(A) RNA were
 prepared, a cDNA library was made, and the cDNA clones
 were in vivo pluscripted to give pluscript phagemids in the
 TJ Close lab (Choi, Close, Fenton) at the University of
 California, Riverside. Plasmid DNA preparations and DNA
 sequencing were performed in the OD Anderson lab (all
 other authors)."
 BASE COUNT 142 a 140 c 150 g 147 t
 ORIGIN

[illegible]

Dbb
61 GTTGGTCACAGAGATTTTGCATCTGCTGAATAACCATGTGGATATCTTGTTGATTTTACACA 120

Oy 420 gagacagacatcgaacgcgctgycagcaagcttgycgttaagaacaactcatgatcttcgag 479

Dbb 121 GATTTTGGTGGCAA---GCTCGTTTCGATTTCCTGAAGGCAAACTTTTGCATCTTGAA 177

Oy 480 gaccgcaagcttcgcctgacatatgtgaaacaggltiaagctcgagtaactcccgagctgtac 539

Dbb 178 GAACCCAAATTCGCTGACATGGGACACCGGTAACTATGCATATTGAAGAAGAAATTTTC 237

Oy 540 cgtalcgcggagctggygcgabatattaccaatgacacagcgcgtcacccgcgccggtgtata 599

Dbb 238 CGCATATCAGACTGGGGCCGATTTGTCTACACGCCGATGAGTACCTCGGACCTGGGAAATATA 297

Oy 600 gccggcgccaagagagcgcgcgaaactggcccacagaagaccagggagttgctgatactgctg 655

Dbb 238 GATGGCTTGAAGCTC-----AAGGTTTTGCCAAAGGAAAAGGTTTACTTTGGTT 348

Oy 660 gcaagcgtctctcttccagagctcttltvgcgcgcgagagactaacgcgcggcgtlctgtgaa 719

Dbb 349 GCTGATGAGAGCGCACCGAACACTTTCATGTAGAGATTACACCAAGCGCCGTGAAG 408

Oy 720 atggcgaaagctggaacgaagtcttgtatcgggttcatcagcg 761

Dbb 409 TTTGCCGAGCAACATTTCGATTTGTTATTCGATTTCTTGTG 450

RESULT	4				
BE446305					
LOCUS					
DEFINITION	BE446305	581 bp	mRNA	EST	25-JUN-2000
	WHE1455_H07_O1325	wheat	etiolated seedling root	normalized cDNA	
	library Triticum aestivum cDNA clone WHE1455_H07_O13			mRNA	
	sequence.				
ACCESSION	BE446305				
VERSION	BE446305.1	GI:9445867			
KEYWORDS	EST				
SOURCE	bread wheat.				
ORGANISM	Triticum aestivum				
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;				

REFERENCE	Magnoliophyta: Liliopsida; Poales; Poaceae; Triticum.
AUTHORS	1 (bases 1 to 581) Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
TITLE	The structure and function of the expressed portion of the wheat genomes - Normalized root cDNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Olin Anderson US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center 800 Buchanan Street, Albany, CA 94710, USA Tel: 51053595773 Fax: 5105595818 Email: oanderson@w.usda.gov Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20 Seq primer: Stragatene SK primer.
FEATURES	Location/Qualifiers
source	1..581 /organism="Triticum aestivum" /cultivar="Chinese Spring" /db_xref="taxon:4565" /clone="WHE145_407.013" /clone_lib="wheat etiolated seedling root normalized cDNA library" /issue_type="Root" /dev_stage="Five day old etiolated seedling" /lab_host="E. coli DH10b" /note="Vector: Lambda uni-zap XR, excised phagemid pBluescript SK; Site.1: EcoRI; Site.2: XhoI; Seeds were surface-sterilized, germinated and grown aseptically in the dark at room temperature on filter paper with water,

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Gossypium hirsutum	Eukaryota: Viridiplantae; Embryophyta: Tracheophyta; Spermatophyta	Magnoliophyta, eudicotyledons: core eudicots; Rosidae; eustroids II	Malvales; Malvaceae; Gossypium.	1 (bases 1 to 597)	
Blewitte, M., Metz, E.C., Davy, D.F. and Burr, B.	ESTs from developing cotton fiber	Unpublished (1999)			
Contact: Ben Burr	Biological Department	Brookhaven National Laboratory	Upton, NY 11973, USA	Tel: 516-344-3396	Fax: 516-344-3407
Email: burr@bnl1.bnl.gov	Seq primer: 73 Primer.	Location/Qualifiers	1..597		
/organism="Gossypium hirsutum"	/cultivar="Acala Maxxa"	/db_xref="taxon:3635"	/clone_lib="Six-day Cotton fiber"	/tissue_type="Immature fiber"	/dev_stage="Six days post anthesis"
/lab_host="XLI-Blue"	/note="Vector: pBluescript II KS"				
BASE COUNT	156 a	117 c	147 g	175 t	2 others
ORIGIN					
Query Match	9.9%	Score 136.4	DB 24	Length 597	
Best Local Similarity	56.2%	Pred. No. 9.6e-29			
Matches	301	Conservative	0	Mismatches 223	Indels 12
					Gaps 2
OY	225	tacgcgaagaagggccacagccacacacatcgcagcttgctagaagctcttcgcatgatg	284		
DB	26	TATCANAANAAGGGCTAAGATTCTTCTAAGATCCAACTGGGAAAAGCTGTTGAGATTAG	85		
OY	285	cacggaagaagaacaacatctctgcgctccctctgaigtgcgagctctagaagctctc	344		
DB	86	GTTAAGAAAGAGATGATCTCTGTTGGGGCTGATGTGGCACATGCTCTGAGTTGCTT	145		
OY	345	gaagctagcagacagcctcggagccgacacatctctcgtcgaagaacatctgcacatctg	404		
DB	146	GATATTTGCGAGAAAGTTGCGACCTGAATTTGCCATTTGMAAGACCCATGTGACATATTC	205		
OY	405	acggagctcgacatcggagcgacagtcgaagccgagcagcagcttcggcgtaagacaac	464		
DB	206	CCTGATTTTCACTCCAGAGATGTCGGTCTTA--GCTTCGGCGCATTTGCCAATAAACCTAAC	262		
OY	465	ttcatgatcttcgagagacccgaagtcgcgtgacatctgcaacagctttaagctgcaglac	524		
DB	263	TTCATGATATTTGMAAGACCGGTAAATTTTGGAGACATCGTGAACAGTAAATATGCAAT	322		
OY	525	tcctcgcgcgtgtaccgtatcgcggagtgaggcgagatataccaatgcacaagcgctcacc	584		
DB	333	GAAAGGGGATCTTCCATATATTTGATGATGGCTGATATTTGTAATGTCACATTAATCTCT	382		
OY	585	ggcccccggtgtgataagccggcggtgaagaaggtcgtgaaacttgccctcaacagaaaccgg	644		
DB	363	GCTCTTGGAATGTTGATGTTGTAATAATTA--GGGTCCTGCTCGTGGTAG	433		
OY	645	gggtctgcagatcgtgcgaagatctctctctccagggctcttgccgcgcggaacataacc	704		
DB	434	GGGCGTGGCTACTTGTCTGAATGAGCTCTGCTGTAACCTTGCCACGGGGAGATACACA	493		
OY	705	gcggagcgtcgttgaacatcggcgaagcttgacgaagaactctgtacggtgcatcagc	760		
DB	494	GCTGCTGCCGTAAATATGCGCAACACACTCCGATTTTGTGCATCGGTTTCATCTC	549		

LOCUS	BE641945	820 bp	mRNA	EST	01-SEP-2000
DEFINITION	Cr12_4_L06_S6	Ceratopteris Spore Library	Ceratopteris richardii		
ACCESSION	CDNA clone Cr12_4_L06 5'	CDNA clone	Ceratopteris richardii		
VERSION	BE641945				
KEYWORDS	BE641945.1	GI:9959615			
SOURCE	EST				
ORGANISM	Ceratopteris richardii				
REFERENCE	Ceratopteris richardii				
AUTHORS	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Filicopsida; Filicales; Pteridaceae; Ceratopteris				
TITLE	Chatterjee, A., San Miguel, P., Stout, S.C., Banks, J. and Roux, S.J.				
JOURNAL	1 (bases 1 to 820)				
COMMENT	Expressed sequence tags of cDNA clones from a C. richardii library				
	Unpublished (2000)				
	Contact: Roux SJ				
	Section of Molecular Cell and Developmental Biology				
	University of Texas				
	Biology Building, Room 16, Austin, TX 78712, USA				
	Tel: 512 471 4238				
	Fax: 512 232 3402				
	Email: sroux@uts.cc.utexas.edu				
	Plate: Cr12_4 row: L column: 06				
	Seq primer: SP6.				
FEATURES	Location/Qualifiers				
Source	1..820				
	/organism="Ceratopteris richardii"				
	/cultivar="Brogan"				
	/db_xref="taxon:49495"				
	/clone="Cr12_4_L06"				
	/clone_lib="Ceratopteris Spore Library"				
	/tissue_type="Gametophyte"				
	/cell_type="Spore"				
	/dev_stage="20 hours after germination initiation"				
	/note="Vector: pcMVSPORT6; EST sequence from cDNA library				
	cdna library constructed from mRNA isolated from C.				
	richardii spores that had developed for 20 hours after				
	their germination had been initiated by white light."				
BASE COUNT	232 a 159 c 219 g 210 t				
ORIGIN					
Query Match	9.5%; Score 131.2; DB 110; Length 820;				
Best Local Similarity	55.6%; Pred. No. 3.4e-27;				
Matches 300; Conservative 0; Mismatches 228; Indels 12; Gaps 2;					
221	atctaacgacagaagaagggccaaagcacacaatctgcagctgtctagaaagcttcgcaat	280			
128	atcatatgatgtgaaagggcacaagcctggcgagacatgccacagcaagcgctgttgacat	187			
281	gatgcagcagagaagaacaacaaatctctgcgctcccttgatgctgcgagcgtctagaagct	340			
188	aatgcagcagcaaaaagagatattctttcactctgctgcagatgtgatttcgtcgtgacct	247			
341	tctggagctagcagacacgcgtcggaacgcacattgtctgctgaagaacacatgtgcacat	400			
248	ccttaccatagctgtaaaagctggcgacctgagattgttgcgaagacacacacgtgcacat	307			
401	actgcagcagacttcgacatcgagacgcacagctcaagcgcgtcgcaagcgttcgagctaa	460			
308	ctatctcctaatttctactcctcctgatTTTTGGTGCAAA---GCGAGCGTGGATTGCGAC	364			
461	caactlcatgatcttcgaagacgcacaagctctcgtacacatttgcacaacacggtttaa	520			
365	caactTTTTTGATTTTGGAGGATAGGAAAGTTCCTGATTATAGGTAACACAGTTCACAT	424			
521	gtactctccgcggtgtgacacgtatctgcgagatggcgagatatccaaatgcacagcgct	580			
425	gattcaaggtgggtcttttcaagattgttagagtggtgcacacacattgtaaaatgcacac	484			
581	cacgggccccgggtgtgatatagccgggtctgaagaagcgttcgaaacacggtctcaaga	640			
485	ttccagggccacaggaattatagatgctttaaactCAA-----GGCGCTGCATTATTA	535			

117: em_estp16:*
 118: em_estp17:*
 119: em_estp18:*
 120: em_estp19:*
 121: em_estp20:*
 122: em_estp21:*
 123: em_estp22:*
 124: em_estp23:*
 125: em_estp24:*
 126: em_estp25:*
 127: em_estp26:*
 128: em_estp27:*
 129: em_estp28:*
 130: em_estp29:*
 131: em_estp30:*
 132: em_estp31:*
 133: em_estp32:*
 134: em_estp33:*
 135: em_estp34:*
 136: em_estp35:*
 137: em_estp36:*
 138: em_estp37:*
 139: em_estp38:*
 140: em_estp39:*
 141: em_estp40:*
 142: em_estp41:*
 143: em_estp42:*
 144: em_estp43:*
 145: em_estp44:*
 146: em_estp45:*
 147: em_estp46:*
 148: em_estp47:*
 149: em_estp48:*
 150: em_estp49:*
 151: em_estp50:*
 152: em_estp51:*
 153: em_estp52:*
 154: em_estp53:*
 155: em_estp54:*
 156: em_estp55:*
 157: em_estp56:*
 158: em_estp57:*
 159: em_estp58:*
 160: em_estp59:*
 161: em_estp60:*
 162: em_estp61:*
 163: em_estp62:*
 164: em_estp63:*
 165: em_estp64:*
 166: em_estp65:*
 167: em_estp66:*
 168: em_estp67:*
 169: em_estp68:*
 170: em_estp69:*
 171: em_estp70:*
 172: em_estp71:*
 173: em_estp72:*
 174: em_estp73:*
 175: em_estp74:*
 176: em_estp75:*
 177: em_estp76:*
 178: em_estp77:*
 179: em_estp78:*
 180: em_estp79:*
 181: em_estp80:*
 182: em_estp81:*
 183: em_estp82:*
 184: em_estp83:*
 185: em_estp84:*
 186: em_estp85:*
 187: em_estp86:*
 188: em_estp87:*
 189: em_estp88:*

190: gb_gss25:*
 191: gb_gss26:*
 192: gb_gss27:*
 193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query length	DB ID	Description
1	136.4	9.9	597 24	AT1731663
2	131.2	9.5	820 110	BE641945
3	128.4	9.3	579 107	BE405348
4	128.4	9.3	581 108	BE446305
5	116	8.4	775 107	BE394063
6	107.6	7.8	557 133	BE019969
7	107.4	7.8	586 94	AW704630
8	107.2	7.8	639 93	AW626116
9	104.8	7.6	465 105	BE210781
10	99.4	7.2	557 12	AA802030
11	97.4	7.1	610 105	BE203967
12	96.2	7.0	461 28	AL369478
13	95.6	6.9	391 3	AA166990
14	95.4	6.9	570 139	BP044448
15	86.4	6.3	987 106	BE300085
16	86.2	6.2	328 31	AV168581
17	85.4	6.2	550 14	AA944506
18	84.2	6.1	420 2	AA128100
19	80.8	5.9	918 109	BE542503
20	79.6	5.8	698 39	AM086516
21	79.2	5.7	549 94	AW755892
22	78.2	5.7	245 40	AM098792
23	75	5.4	649 40	AW116069
24	74.6	5.4	431 2	AA125810
25	74.4	5.4	724 96	AM905992
26	74.2	5.4	668 138	BE951337
27	70.8	5.1	591 89	AM360558
28	67	4.9	247 5	AA328585
29	66.8	4.8	719 26	AI894226
30	65	4.7	148 96	AM851491
31	63.2	4.6	630 8	AA537289
32	60.6	4.4	117 96	AM851496
33	60	4.3	777 140	C97586
34	59.8	4.3	809 135	BE745374
35	59.6	4.3	153 96	AM851587
36	58.8	4.3	617 23	AI655439
37	57.8	4.2	516 93	AM659213
38	57.8	4.2	564 23	AI668476
39	56.8	4.1	279 5	AA295675
40	56.8	4.1	529 22	AI597621
41	55.8	4.0	276 6	AA384977
42	55.2	4.0	636 26	AI862776
43	54.2	3.9	939 108	BE470313
44	54.2	3.9	546 19	AI326762
45	53.8	3.9	498 13	AA919562

ALIGNMENTS

RESULT 1
 LOCUS AT1731663 597 bp mRNA EST 11-JUN-1999
 DEFINITION BNLGH110421 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar
 to URIDINE 5'-MONOPHOSPHATE SYNTHASE (UMP, mRNA sequence).
 ACCESSION AT1731663
 VERSION AT1731663.1 GI:5050515
 KEYWORDS EST.
 SOURCE upland cotton.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2001, 20:11:13 ; Search time 1348.49 Seconds
(without alignments)
7171.231 Million cell updates/sec

Title: US-09-582-779A-1
Perfect score: 1380
Sequence: 1 ctcgagcactcattggaag.....aataagatgcgtgaattc 1380

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estbda:*
42: em_estfun:*
43: em_estchum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estln1:*
64: em_estln2:*
65: em_estln3:*
66: em_estln4:*
67: em_estlov1:*
68: em_estlov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estrol1:*
75: em_estrol2:*
76: em_estrol3:*
77: em_estrol4:*
78: em_estrol5:*
79: em_estrol6:*
80: em_estrol7:*
81: em_estrol8:*
82: em_estrol9:*
83: em_estrol10:*
84: em_estrol11:*
85: em_estrol12:*
86: em_estrol13:*
87: gb_esta1:*
88: gb_esta2:*
89: gb_esta3:*
90: gb_esta4:*
91: gb_esta5:*
92: gb_esta6:*
93: gb_esta7:*
94: gb_esta8:*
95: gb_esta9:*
96: gb_esta10:*
97: gb_esta11:*
98: gb_esta12:*
99: gb_esta13:*
100: gb_esta14:*
101: gb_esta15:*
102: gb_esta16:*
103: gb_esta17:*
104: gb_esta18:*
105: gb_esta19:*
106: gb_esta20:*
107: gb_esta21:*
108: gb_esta22:*
109: gb_esta23:*
110: gb_esta24:*
111: gb_esta25:*
112: em_estchum21:*
113: em_estchum22:*
114: em_estchum23:*
115: em_estcom1:*
116: em_estcom2:*

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: Apr11 3, 2001, 21:34:18 ; Search time 2742.82 Seconds
(without alignments)
2574.903 Million cell updates/sec

Title: US-09-582-779a-1

Perfect score: 1380
Sequence: 1 ctcgagcactcattggaag.....aatatgagtcgcttgatc 1380

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database : GenEmbl :
1: gb_ba1 :
2: gb_ba2 :
3: gb_om :
4: gb_om :
5: gb_ph :
6: gb_ph :
7: gb_ph :
8: gb_ph :
9: gb_ph :
10: gb_ph :
11: gb_ph :
12: gb_ph :
13: gb_ph :
14: gb_ph :
15: gb_ph :
16: gb_ph :
17: gb_ph :
18: gb_ph :
19: gb_ph :
20: gb_ph :
21: gb_ph :
22: gb_ph :
23: gb_ph :
24: gb_ph :
25: gb_ph :
26: gb_ph :
27: gb_ph :
28: gb_ph :
29: gb_ph :
30: gb_ph :
31: gb_ph :
32: gb_ph :
33: gb_ph :
34: gb_ph :
35: gb_ph :
36: gb_ph :
37: gb_ph :
38: gb_ph :
39: gb_ph :
40: gb_ph :
41: gb_ph :
42: gb_ph :
43: gb_ph :

44: em_hcg2 :
45: em_hcg3 :
46: em_hcg4 :
47: gb_ph :
48: gb_ph :
49: gb_ph :
50: gb_ph :
51: gb_ph :
52: gb_ph :
53: gb_ph :
54: gb_ph :
55: gb_ph :
56: gb_ph :
57: gb_ph :
58: gb_ph :
59: gb_ph :
60: gb_ph :
61: gb_ph :
62: gb_ph :
63: gb_ph :
64: gb_ph :
65: gb_ph :
66: gb_ph :
67: gb_ph :
68: gb_ph :
69: gb_ph :
70: gb_ph :
71: gb_ph :
72: gb_ph :
73: gb_ph :
74: gb_ph :
75: gb_ph :
76: gb_ph :
77: gb_ph :
78: gb_ph :
79: gb_ph :
80: gb_ph :
81: gb_ph :
82: gb_ph :
83: gb_ph :
84: gb_ph :
85: gb_ph :

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1380	100.0	1380 81 A94699	A94699 Sequence 1
2	397.4	28.8	1464 6 AB021640	AB021640 Saccharom
3	391	28.3	1021 47 YSACANDIDA	LI361 Candida gla
4	372.8	27.0	1167 47 KLURA3	Y00454 Kluyveromyc
5	372.8	27.0	1764 47 YSKURA3A	D00431 Yeast Oroti
6	370.4	26.8	2330 21 E11619	E11619 Candida uti
7	370.4	26.8	2330 81 AR065318	AR065318 Sequence
8	369.6	26.8	1469 47 KMURA3GN	221934 K.marxianus
9	368.6	26.8	1469 81 A39345	A39345 Sequence 11
10	368.6	26.8	3750 7 CRU279021	AU279021 Candida r
11	359.2	26.0	1166 47 YSCODCF	K02207 Yeast (S.ce
12	359.2	26.0	4133 12 AF106619	AF106619 Cloning v
13	359.2	26.0	4342 12 SCU54830	U54830 Synthetic E
14	359.2	26.0	6107 12 SCU54829	U54829 Synthetic E
15	359.2	26.0	6119 12 SCU54828	U54828 Synthetic E
16	359.2	26.0	9344 12 SYNPCS19X	M98046 PCS19 Cloni
17	359.2	26.0	62643 47 SCE5871	U18530 Saccharomyc
18	359.2	26.0	115355 84 CEY55D9	AL008876 Caenorhab
19	357.6	25.9	798 7 AF279259	AF279259 Zygosacch
20	357.6	25.9	5225 12 CVU63018	U63018 Cloning vec
21	357	25.9	4303 12 SYNIP352	LI4759 YIP352, yea

22	357	25.9	5056	12	SCU64694	U64694	Yeast Integ
23	357	25.9	5181	12	SYNEP352	L14758	YEP352, yea
24	357	25.9	7964	12	CVU28790	U28790	Cloning vec
25	357	25.9	11741	12	AF114752	AF114752	Shuttle v
26	356	25.8	1101	13	AX033928	AX033928	Sequence
27	356	25.8	1170	43	YSCOPCD	K02206	Yeast (S.ce
28	356	25.8	3797	12	CYR1P1211	X75452	Saccharomyc
29	356	25.8	3799	12	CVPL34	X70479	Integrative
30	356	25.8	4102	81	AB4775	AB4775	Sequence 20
31	356	25.8	4164	81	AR062872	AR062872	Sequence
32	356	25.8	4164	81	AR080476	AR080476	Sequence
33	356	25.8	4164	81	169323	169323	Sequence 1
34	356	25.8	4319	12	CVOPFL4S	X70266	multicopy S
35	356	25.8	4373	12	PRS306	U03438	Yeast Integ
36	356	25.8	4384	12	PRS406	U03446	Yeast Integ
37	356	25.8	4610	12	CVPL38	X70482	replicative
38	356	25.8	4702	12	XU34314	U74314	E.coli/K1uy
39	356	25.8	4763	12	CVPL64M	X70273	single stra
40	356	25.8	4763	12	CVPL64P	X70274	single stra
41	356	25.8	4887	12	PRS316	U03442	Yeast cent
42	356	25.8	4898	12	PRS416	U03450	Yeast cent
43	356	25.8	4933	81	AR062873	AR062873	Sequence
44	356	25.8	4933	81	AR080477	AR080477	Sequence
45	356	25.8	4933	81	169324	169324	Sequence 2

ALIGNMENTS

RESULT 1
LOCUS A94699 1380 bp DNA
DEFINITION Sequence 1 from Patent WO9336432.
ACCESSION A94699
VERSION A94699.1 GI:6778966
KEYWORDS
SOURCE Eremothecium gossypii.
ORGANISM Eremothecium gossypii.
Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
Eremotheciaaceae; Eremothecium.

REFERENCE
Pompejus, M. and Santos, G.M.
1 (bases 1 to 1380)
OROTIDINE 5'-PHOSPHATE DECARBOXYLASE-GENE, GENE CONSTRUCT
CONTAINING SAID GENE AND THE UTILIZATION THEREOF
Patent: WO 9336432-A 22-JUL-1999.
JOURNAL BASF AG (DE); POMPEJUS MARKUS (DE)
FEATURES
source
Location/Qualifiers
1..1380
/organism="Eremothecium gossypii"
/db_xref="taxon:33169"
/clone="URA3"
1..199
210..1013
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB70440.1"
/db_xref="GI:6778966"
/translation="WSTSVARAKAHNSPVARKLLALMHEKTNLCASLDVRRSLK
LEIADPTLPHICLTKTHVDILFDIETTVKPLQDLAKNHMFIEDKRFIDIGTVK
LOYSSGVYRIAMADITNAGVTGPGVAGLAKAEAKLSQEPRLGLMLAEISQSLA
RBDYAGVEMAKLDEDFYVIGFIAQRDNGRADGDMILMPGVLDKGGGLGQOYR
TYDEVVSDTDVITIGRGFLFDKGRDPKVGARVYKAGAEATLRNGERS"

3'UTR
BASE COUNT 334 a 354 c 375 g 317 t
ORIGIN

Query Match 100.0%; Score 1380; DB 81; Length 1380;
Best Local Similarity 100.0%; Pred. No. 3.6e-295;
Matches 1380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ctgagcaactcttggaagccctgcgaagacctatatactgctcaagttccta 60
|||||

Db	1	CTCAGCAACTCANTTGAAGCCCTTGCNAAGACCTCTATATCTGCTCAAGTTCTGA 60
OY	61	ctatcatgtatgtctgtactacagaaaattttgtctatagcttgcaagaacacata 120
Db	61	CTATCATGTATGTCTGTCTACTACAGAAAATTTTGTCTATAGCTGCNAAGACATCA 120
OY	121	catatcttcgatgtgtgtgaagctccacatacagaagtaagcttgtttaagctgataca 180
Db	121	CATCATCTTGATGTGTGTGAGCTTCACATACAGTAAGCTTTGTATAGCTGATACACA 180
OY	181	taaggttctaccagaccagccatgtccacatagtcaagaaattctacgcagaaggcca 240
Db	181	TAGGTGTCTACCGACCTAGCCATTCGCATGTCCACATGTCCAAAGTAATCTTACGCGAAGAGGCCA 240
OY	241	aggcacaaatttgcacagcttgctagaagcttctgacatgtgacagcaggaagaaacca 300
Db	241	AGGCACCAATTTCCCGAGTTGCTAGAAAAGTTCTGGCATTTGATGACAGCAACAAAACCA 300
OY	301	atctctgcgcttcccttgatgtgctgagcagctctagaagcttctgagctagcagacgc 360
Db	301	ATCTCTGCGCTTCCCTTGATGTGTGAGCTGTGAGAAAGCTTCTGAGACTAGCAGACACGC 360
OY	361	tgagaccgcacatttgcctgctgtaagaacacatgtgcacatactgacagagacttcgacatcg 420
Db	361	TGGACCGCACATTTGTCTGCTGAAGACACATGTGCACATCTGACGAGACTTGCACATCG 420
OY	421	agaagcagttcaagccgctgagagagcttgcggttaagcacaacttcatatcttcgagg 480
Db	421	AGAGCAGCTCAAGCCGCTGAGAGAGCTTGGGCTTAAGCAACATCTATATCTTCAGAG 480
OY	481	accgcaagctcgcctgacacattgacacacagcttlaagctgcagtaactctccgcggtgtacc 540
Db	481	ACCGCAAGTTGCTGACATTTGGCAACAGGTTAAGCTGACGTACTCTCCGCGGTGACC 540
OY	541	gtatcgcgagtgaggcgagatactacaaatgcacagcgcttcaacgcgcccgggtgtgatag 600
Db	541	GTATCGCGAGTGGCGGATTTACCAATGCACAGCGCTTACCGCGCGGCTGTGATAG 600
OY	601	ccggagctgaaggaagcttgagaaacttgccctacacaggaagcccaagggtgtctgtatgctg 660
Db	601	CCGGAGCTGAAGGAAGCTTGCAAACTTGCCCTACAGGAACCCAGGGGTTCTGATGCTGG 660
OY	661	caagctctctctcacaagagctcttggcgcgcgagagacataccgcyggcgctgctgtgaaa 720
Db	661	CAGAGCTCTTCTCAGAGGCTCTTTGGGCGCGGAGACTATACCGCGGCGGTGAAA 720
OY	721	tggcgagctgagcgaagagacttgtgatcggtgttcaatcgcgcgagcggtgaatgggtgggc 780
Db	721	TGGCGAAGCTTGAGCAAGACCTTTGTGATCGGGTTCAATCGCGCACAGCTGACATGGGTGGGC 780
OY	781	gtgagagagcttgaactgagctgtacatacagcccgggggtttggccttggaagcaaaaggag 840
Db	781	GTGAGAGAGCTTTGACTGTGCTCATATGACCCCGGGGTTGGCTTGAGACAGCAAAAGAG 840
OY	841	acggcttgagccagcagctacccagctggaatgaatgagcttgaagcagctgacagatgtga 900
Db	841	ACGGCTTGAGCCAGCAGATGACCGAGGGATGAGTGTGTCAGAGCAGATACGATGTGA 900
OY	901	tcaatgttggcagagagctcttgaagaagggaagaccccaaggttcgaaggtgtgcccgt 960
Db	901	TCAATGTTGGCAGAGGCTCTTTGACAAAGGGAAGACCCCAAGGTGAGAGGTGCCCGCT 960
OY	961	accgcaagggcggtttggaagcttacttggcggtatgaggcgagagacttcgtaagctatacg 1020
Db	961	ACCGCAAGGCGGTTGGAGGCTTACTTGGCCCTATAGGGGAGACTTCTGATGCTATCG 1020
OY	1021	ctggcgccacagatataagcggatccacccgcgataccaatccacagaaacttttg 1080
Db	1021	CTGGCGCCACAGATATAGCGGATTCACCGCGGATTCACATCTCAGCAACTTTTGG 1080
OY	1081	taattataatgcccattgtcccttatttccgagcttggctgcgggatacggtttataagcag 1140
Db	1081	TAAATATATGCCCTATTTGCCCTTATTTCCAGAGCTGTGCGCGGATGCGTTATAGACGC 1140

QY 1141 gcaacaaatgtatcttctgtctagatgacatccacacatctgcaagcttggtgtgtg 1200
|||||
DB 1141 GCAACAAAGTGTACTTGTTCAGTAGCATCATCAAACTTGTGAGGTGGGTGG 1200
QY 1201 aagccctgcgcggagataatctgattacccgcaatctggaatctgtttatgaaa 1260
|||||
DB 1201 AAGCCCTGCGCGGATTAATTCGTATTACCGCCTTGTGAATTTGTTATGAAA 1260
QY 1261 actctcaacttggagctactagagcacaatactcgacacaaagccctccctacacactg 1320
|||||
DB 1261 ATCTTCACTTGGGCTTAAGTACAGCCATTAAGTACAGCCCTTCTTACACACTTGC 1320
QY 1321 agctgggaactaaagtgacacgaatagcaataatagcaataatgagctgtgaattc 1380
|||||
DB 1321 AGCTGGAGTAAGTACAGCAAGTACCAATTAATTAAGCAATTAATGATGATGATTC 1380

RESULT 2
LOCUS AB021640 1464 bp DNA PLN 06-APR-1999
DEFINITION Saccharomyces exiguus SeURA4 gene for orotidine-5'-phosphate decarboxylase, complete cds.
ACCESSION AB021640
VERSION AB021640.1 GI:4589378
KEYWORDS orotidine-5'-phosphate decarboxylase.
SOURCE Saccharomyces exiguus (strain:Yp74L-3).
ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (sites)
HISATOMI,T., KODAMA,T., TOBE,H., MOTO-OKA,M., WATANABE,A. and TSUBOI,M.
Molecular cloning and DNA analysis of the orotidine-5'-phosphate decarboxylase gene from the yeast Saccharomyces exiguus Yp74L-3 (bases 1 to 1464)
JOURNAL Direct Submission
AUTHORS Submitted (18-DEC-1998) to the DDBJ/EMBL/Genbank databases. Taisuke HISATOMI, Fukuyama University, Faculty of Engineering, Department of Biotechnology; Gakken-cho 1, Fukuyama, Hiroshima 729-0292, Japan (E-mail:hisatomit@f.fu.fukuyama-u.ac.jp, Tel:+81-849-36-2111(ex 4627), Fax:+81-849-36-2459)
JOURNAL Location/Qualifiers
TITLE 1..1464
FEATURES
source
1..1464
/organism="Saccharomyces exiguus"
/strain="Yp74L-3"
/db_xref="taxon:34358"
488..1294
/gene="SeURA4"
488..1294
/gene="SeURA4"
/product="orotidine-5'-phosphate decarboxylase"
/protein_id="BAA6736.1"
/db_xref="GI:4589378"
/translation="MAFATYERASVHSPVASKLLRLIMASKRTNLCSLDVRYTAELLSLVQVGYICLTATHVDLDDFSMEGVKPLKELAAKYNLLEDEKRFADIGNTVKLQYSSGVYKIVQWADITNAGVTGSGIYGLKEAQNDSDEKGLIMLAESSKSLAHCEYTGVEIAKTDKQFVFGFIAGRMGGRREGFDWIMTPGVGLDKGADLGGQYRTVDVYVSGSDIILIVGRGLFGKGRDPTAESRRYRAGWDAYIKRCNAASN"

BASE COUNT 367 a 337 c 445 g 315 t
ORIGIN

Query Match 28.8%; Score 397.4; DB 6; Length 1464;
Best local Similarity 69.9%; Pred. No. 8.1e-78;
Matches 536; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

QY 223 cttacgcagaagaagcacaacatcgccagctgtctagaagctcttgcatga 282
|||||
DB 501 CGTACGAGAGAGACGCTGGTGCACGCGAGTCCCTTCCCTGGAAGCTGCGGCGATCA 560

QY 283 tgcacgagaagaacaaatctctgcctcccttgatgtgcggaagctctagaagcttc 342
|||||
DB 561 TGCCCTCCCAAGAGACGAACTGTGTGCGCTTGCTTACGTGCGACGACTGGCGAGTGC 620
QY 343 tggagctagacagacacgccttgagccgacatctgtctgtgaagacacatgctgacatac 402
|||||
DB 621 TCTCGCTGGTGACCAAGTGCGTCCCTACATTTGTCTTTGAAGACTATGTGGACATCC 680
QY 403 tgcgcgaacttcgacatcgcgagacagctaaagcgcctgcgaagctctgcgttaagcaca 462
|||||
DB 681 TGGACATTTCTCGATGAGGAGGAGCGTGAACCCCTTGAAGAGCGTTGCTGCAAGTACA 740
QY 463 actctatgactctcgcgagccgcaagctgcgtgacatctgcaacatgcaacggttaagctgagc 522
|||||
DB 741 ACTTCTTGAATCTTCGAGACACCGGAATTTGCGGACATCGGGAACACGCTGAAGTTCAC 800
QY 523 actcctcgcgcgttgatccgctatcgcgagctgagcgagatactcaatgacacagcgctgca 582
|||||
DB 801 ACTCTCTCCGCTGTGTACAGATCGTGCAGTGGGCGACATCAGAAACGCGACGCTGCA 860
QY 583 ccgagcccggttgatagccgagctggaagagctgcggaactgcctcaccaggaacca 642
|||||
DB 861 CCGGTAGTGGATTTGAAAGGTTGAAAGAGCCGCCAGAGGAACAGTGAACGACCGC 920
QY 643 ggggggtgctgctgctgctgagagctctctcagagctctcttgccgagcgagactata 702
|||||
DB 921 GTGGGTTCTGATGCTGTGGGAATTTGCTCGAAGAGTTTCACTTGCACACGGGAGTACA 980
QY 703 ccgagcgagctgtgaaatgagcgaagctgagcgaagacttgatcgggtatcagcgac 762
|||||
DB 981 CCACGTGGCACTGTAGAGATTTGCAAAACGACACACAGTGTGTCTGGGTTCATTTGCGC 1040
QY 763 agcgatgacatgggtgagcgctgcagacgcttgactgctacatgagcccgaggtg 822
|||||
DB 1041 AGAGGGCATGTGGCGGAGACAGCAAGTTTCTGACTGCTGATCAACGCCGCTGTTG 1100
QY 823 gctctgacgacaaagagagcgcgcttgagccagctgacgcgctgagatgagctgca 882
|||||
DB 1101 GCGTGACGATTAAGGCGCATGTCCTTGACACACACTACAGGACCGTGGACGATGCTGT 1160
QY 883 ggaagagctcgaatgacatctgttgcaagagagctcttgcaagaggaagaccca 942
|||||
DB 1161 CCACGGGCTCGGACATCATCATTTGCGGACAGAGCATTTGGCAAGGCGACGCCAA 1220
QY 943 agctcgaaggtgctccgctacccgcaagcggttgagagcttactg 989
|||||
DB 1221 CCGCCGAGAGACGACGCTACAGAGACGCTGGGTGGGACCGCTACTTG 1267

RESULT 3
LOCUS YSACANDIDA 1021 bp DNA PLN 27-JUL-1994
DEFINITION Candida glabrata orotidine-5'-phosphate decarboxylase gene, complete cds.
ACCESSION L13661
VERSION L13661.1 GI:388364
KEYWORDS orotidine-5'-phosphate decarboxylase.
SOURCE Candida glabrata (strain 85/038) DNA.
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales; anamorphic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1021)
Zhou,P., Szczypka,M.S., Young,R.J. and Thiele,D.J.
A system for gene cloning and manipulation in the yeast Candida glabrata
JOURNAL Gene 142, 135-140 (1994)
MEDLINE 94237479
FEATURES
source
1..1021
Location/Qualifiers
/organism="Candida glabrata"
/strain="85/038"
/db_xref="taxon:5478"

QY 563 taccaatgcaacacgacgacacgacccggtgtgtatagccggcgtgaagaagcgtcgaa 622
 Db 531 CACCAACGCCCGACGGGCTACTGCTGCTGATGTGCTGGCTTGAACAAGGTCCGCA 590
 QY 623 actgacctcaagaacccagcggtcgtcgtatgctcgcaagctctctcaaggtc 602
 Db 591 AGAGGTACCAAGAACCAAGGCGATTATTGATGCTGTGTAATTTCTTCCAAAGGGTTC 650
 QY 683 ttggcgcgaggaacatacccgcgctcgttgaatgtcgaaagcttgagcaagaatc 742
 Db 651 TCTAGCACACGGTGAATATACCTAAGGCTACCGTGTATATGCAAGAAGTATTAAGTTT 710
 QY 743 tgtatcgtggttcaacgcgacgacgtgacaatggtgtggtcgacgacgtcttactgct 802
 Db 711 CGTATTGGGCTTCTCTCTAGAAAGATATGGGAGAAAGAAAGAGGCTTTATTTGGCT 770
 QY 803 catcatgaccccgaggtctggccctgacgacaaagagacgagccctggcgacgacgtacg 862
 Db 771 AATCATGACCCCAAGCTTACGTTTACACGACAAAGGCGATTCATTTGGCTACGATACAG 830
 QY 863 caaggtgatatgaggtcgacgacgacgtacgacgtgacatgtgtgtcgacagaggtctc 922
 Db 831 AACCGTCGACGAAGTTTGAAGTGTGATCATCATCATCTTGTGGCAGAGGACTTTT 890
 QY 923 tgcacaggaagagaccccaaggtcgaggtgtgccgtacgcgacagcggttggagagc 982
 Db 891 CGCCAAAGCGTAGACATCTTAAGTTGAAGTGAAAGATACGAATTCGTGATGGGAAGC 950
 QY 983 ttac 986
 Db 951 GTAC 954
 RESULT 5
 LOCUS YSKURA3A 1764 bp DNA PLN 02-FEB-1999
 DEFINITION Yeast orotidine 5'-phosphate decarboxylase (URA3) gene, complete cds.
 ACCSSION D00431.1 GI:218526
 VERSION D00431.1
 KEYWORDS URA3; orotidine-5'-phosphate decarboxylase.
 SOURCE Kluyveromyces laetis (strain IF01267) chromosomal DNA, clones PM1214.5).
 ORGANISM Kluyveromyces laetis
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 REFERENCE 1 (bases 1 to 1764)
 AUTHORS Mizukami, M. and Hishinuma, F.
 TITLE Isolation and nucleotide sequence analysis of the URA3 (orotidine 5'-phosphate decarboxylase) gene of Kluyveromyces laetis
 JOURNAL Agric. Biol. Chem. 52, 3067-3071 (1988)
 COMMENT Transcription of URA3 of K. laetis begins around -30 to -40 from the initiation ATG codon. It was identified by SI mapping.
 FEATURES
 source
 1. 1764
 /organism="Kluyveromyces laetis"
 /db_xref="taxon:28985"
 TATA_signal
 207..214
 /note="putative"
 TATA_signal
 316..320
 /note="putative"
 CDS
 499..1302
 /note="orotidine 5'-phosphate decarboxylase"
 /codon_start=1
 /protein_id="BAA00333.1"
 /db_xref="GI:218527"
 /translation="MSTKSYTSAETHASPVASKLLRLMDEKTNLCAISLVRSDEL
 LKIVETLPTICLLKTHVDILDFSTEGVYVPLKLAIEYKFLIFEDRFADIGNTVK
 LQYTSQVRIASMSDITNAHGTAGIVAGLKGAGAEVTKERGLMLAEISSKSLA
 HGEYKTVDIASDKDFVITGFIAQNDMGREGHFDWLMTPGVGLIDKGDALGQOVR
 TVDEVVSGSDIILIVGRGLFAKGRDPKVEGERYRNAAGEAAYOKRISABH"
 BASE COUNT 553 a 327 c 385 g 499 t
 ORIGIN 284 bp upstream of XhoI site.

Query Match 27.0%; Score 372.8; DB 47; Length 1764;
 Best Local Similarity 67.2%; Pred No. 2.2e-72;
 Matches 527; Conservative 0; Mismatches 237; Indels 0; Gaps 0;
 QY 203 ttgcacatgltcaacgaaatcttaccgacaagaagggccaagacaaatcttcgacgttgc 262
 Db 492 TTGCACCATGTCCACAAATATATACAGTGAAGGTGAGACTCATGCAAGTCCGGTTC 551
 QY 263 tgaagaagcttcggacatgtagcagagaagaacaaatctctcgtcgtctcccttgatgt 322
 Db 552 ATCGAAACCTTTACGTTTAAATGATGAAAAGAACCAATTTGTGTGCTTCTCTGAGCT 611
 QY 323 gcgagcgtcagaagaagctcttgagcttagcagacgcgtcgagacgcgacatgtctgct 382
 Db 612 TCGTTGACATGATGACCTTTGAACCTTGTAAACGTTTGGTCCATACATTTGGCTTTT 671
 QY 383 gaagacacatgtcgaacatactgacgacttcgacatcgaagacgaagtcgaagcgtctga 442
 Db 672 GAAACACACAGTTGATATCTTGATGATTTTCAGTTATGAGGGTACTGTCGTTCCATTGAA 731
 QY 443 gcaagcttggcgttaagcacaactcatgactcttgagagacgcgaagctcgtcgaatcgt 502
 Db 732 AGCATTTGACAGAAATTCAGATTTCTTGAATTTTGAAGACAGCAAAATTCGCCGATATCG 791
 QY 503 caaacggttaagctgcaatctactcctcgcggtgtagcgtacgcgaggtgagcgagatc 562
 Db 792 TTAACAGTCAATATTAATATATATACATCGGCGCTTTACCTATCGCAGATGCTGATAT 851
 QY 563 taccaatgcaacacgacgacacgaccccggtgtgtatagccggcgtgaagaagcgtcgaa 622
 Db 852 CACCAACGCCCGACGGGCTTACTGCTGTGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 911
 QY 623 actgacctcaagaacccagcggttgcgtgactgctgacgagacgtctctcgaaggtc 682
 Db 912 AAGGTCAACCAAGAACCAAGGCGATTATGATGCTTCTGTAATTTCTTCCAAAGGCTTC 971
 QY 683 ttggcgcgaggaacatacccgcggtcgctgtgaatgtgcaagcgtgagcgaagactc 742
 Db 972 TCTAGACACGCGTGAATATACATAGGCTACCGTTGATATGCAAGAGATGAAGATTT 1031
 QY 743 tgtatcgtggttcaacgcgacgacgtgacgtgtggtgtgagcgtgacgacgtcttactgct 802
 Db 1032 CCTTATTGGCTTCAATTTGCTCAGACAGATATGAGGAGAGAGAAAGGCTTTGATGGCT 1091
 QY 803 catcatgaccccgaggtgtgtgcttgcgtgacgacaaagagacggtccgtgacgacgtacg 862
 Db 1092 AATCATGACCCCGACGTTGATGTTAGACGACAAAGCGCATGCTGTCAGCAATACAG 1151
 QY 863 caaggtgatatgaggtcgacgacgacgtgacgtgtgacatgtgttgacgagaggtcctc 922
 Db 1152 AACCGTCACGAAAGTTGTAAGTGTGATGATCATATTCATCTTGTGGCAAGGACTTTT 1211
 QY 923 tgacaaaggaagaccccaaggtcgaggtgtgcccgtacacgcgacagcggttggagagc 982
 Db 1212 CCCCAAGGCTAGAGATCTTAAGTTGAAGTGAAGATACGAATATGCTGATGGGAAGC 1271
 QY 983 ttac 986
 Db 1272 GTAC 1275
 RESULT 6
 E11619
 ID E11619 standard; DNA; FUN; 2330 BP.
 AC E11619;
 XX
 SV E11619.1
 XX
 DT 08-OCT-1997 (Rel. 52, Created)
 DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)

[illegible]

Db	1626	ACGGTGTACCCGGTGCAGGTATGCTCAAGGGCTTAAAGAGGCTCCACAGAAACACAGG	1685
Qy	634	aggaaccccaaggagggtctgctgcatgctgagagagctctctcaagggtccttgcgcgcg	693
Db	1686	ATGACCAAGAGAGGGCTGTGTGATGCTTTGGGAGACATGAGCTCCAAAGGCTCTTTGGCCACG	1745
Qy	694	gagactataccgcgggctgctgttgaatgvgcgaaagctgagaaagacttttgatccggct	753
Db	1746	GGACATATATACCCAGAGAGACCGTGGAGATTGCCAAAACTGATATAGACATTTTGTATTGGAT	1805
Qy	754	tcatcgcgcaagctgacatgvgctggcgctgagacagctcttactgctgctcaltcaatgaccc	813
Db	1806	TCATGCGCACAGAGACATGAGGGTGGCCAGAGAAATGGGTTCACATGATCATATGATGACAC	1865
Qy	814	cgggggcttgcccttgagcagcaaaagagacggtcttggccagcatgacagcagctgtagtg	873
Db	1866	CAGCGCTGGAGATCCAGCATATAGGGCCGACTCCCTGGGCCCAACATGATCAACACTGTTCGATG	1925
Qy	874	aggttgctcagcgcaagctgaccgaatgcatatgctgttggcgaagaggtctcttggcaaggaa	933
Db	1926	AGGTTGTGAGTGGTGGCTTGTGACATCATCATCTGTGTGTGATGAGGCTTGTGTGAAAGGGA	1985
Qy	934	ggagcccaaggctcgagggctgccctacccgcaaggccggttgggagagctta	985
Db	1986	GAGATCCAAACATGGAAGGTGAGCGTTATATAGAAAGACAGGCTGGGATGCTTGA	2037
RESULT	7		
AR065318	AR065318	2330 bp	DNA
LOCUS	DEFINITION	Sequence 3 from patent US 5849524.	PAT
AR065318	AR065318	GI:5995534	29-SEP-1999
KEYWORDS	UNKNOWN.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 2330)		
AUTHORS	Kondo, K., Kajiwara, S., and Misawa, N.		
TITLE	Transformation systems for the yeast candida utilis and the		
JOURNAL	expression of heterologous genes therewith		
FEATURES	Patent: US 5849524-A 3 15-DEC-1998;		
SOURCE	location/Qualifiers		
BASE COUNT	669 a	498 c	613 g
ORIGIN	1..2330	/organism="unknown"	550 t
Query Match	26.8%;	Score 370.4;	DB 81;
Best Local Similarity	67.5%;	Pred. No. 7.1e-72;	Length 2330;
Matches 521;	Conservative	0;	Mismatches 251;
		Indels	0;
		Gaps	0;
Qy	214	caacgaatcttaccgagaaagggccaagcacacaaatctgcacagttgctagaagaagcttc	273
Db	1266	CCACGTTATCGTACACAGAGAGGGCATTCGACAGACCCCTTGGCACTTGGCTAAGCGTGTG	1325
Qy	274	tggcatctgagcaggaagaacaacatctctggtctcccttgatgtgcygagtgctca	333
Db	1326	TTTCGCTTATGAGTCCAAAGAACACACCTGTGTCCAGTGTGATGCTTTCGATCCACAG	1385
Qy	334	gaaagctcttgagctcagcagacagcgtctgggacccgacatctgtctcgtgaagaacatg	393
Db	1386	AGGAGTTGCTCAAGCTCGTTGATACGCTTGTCTTATATCTGTCTGTTGAAGACGCATA	1445
Qy	394	tgcacactgagcagctcagacatgagcagacagctcaagccgcgtcgaagcagctgcg	453
Db	1446	TTGATATCATTTGATGATCTTCTCATGAGTCTACTGTGGCTCCACCTGTTGGACCTTTCAA	1505
Qy	454	ctaagacacactcagcatcttcaggaagcgaagtcgctgacatctggcaacacgtaga	513
Db	1506	AGAGACAAATTTCCATCTTTGAGAGACCGTAAGTTGCTGATATGGCAACACCGTCA	1565


```

repeat_unit      1..38
                  /rpt_type="terminal
                  93..126
                  /note="loxR site"
                  /organism="synthetic construct"
                  135..851
gene             /gene="GFP"
                  /gene="GFP"
                  /codon_start=1
                  /transl_table=11
                  /product="green fluorescent protein"
                  /protein_id="AAB01173.1"
                  /db_xref="GI:1335938"
                  /transl_table="MSKGEELFTGVVPIVELDGVNGHKSFSVGESEGDATYGLTL
                  KFICTGKGLPVPMPPLVTTPFSYGVCFSRYPDHMKOHDFEKSMPGEGYQERTIFPKD
                  DGNYKTRAEVKFEGDPLVNRIELKIDPKEDGILGHLKLEYNNSHNIVIMADKOKG
                  IKVNFIRHNIEDGSYVLIADHYQONTPIGDVPLLPDNIHLTQSLSLSDPNKRDHM
                  VLEFPTAAGITTHGNDELTK"
                  1688..2491
gene             /gene="URA3"
                  1688..2491
                  /gene="URA3"
                  /note="ura3p from S. cerevisiae"
                  /codon_start=1
                  /transl_table=11
                  /product="OMP decarboxylase"
                  /protein_id="AAB01174.1"
                  /db_xref="GI:1335939"
                  /transl_table="MSKATYKKEKRAHPSPVAKLFNIMHEKOTNCASLDIPTYKEL
                  LELVEALGPICLTKHVDTLDEPDSMEGYKPKALSAYNNELFEDRKFADIGNTVK
                  LQYSAYVRIAEKADITNHAHVYGVGIVGLKQAAEEVTKERGLMLLELCKGSLA
                  TGEYTGVDIAKSDKDFVIGFTIARDKMGREGDYMLIMTPGVGLDDGDALGOOYR
                  TVDDVYSTSDIIVGRGLFAKGRDAKVEGERYRKAGWEAYLRRCGOQN"
                  2684..3874
gene             /gene="tet"
                  2684..3874
                  /gene="tet"
                  /note="tet"
                  /codon_start=1
                  /transl_table=11
                  /product="tetracycline efflux protein"
                  /protein_id="AAB01175.1"
                  /db_xref="GI:1335940"
                  /transl_table="MKSNNALVIYICTVLDVAGTCLVNPVPLGLIRDIIVHSDTASH
                  YGVLLALMDKPLCAPVGLASDRFRPVPVLASLGATIDYALIMATTPVLMITVAG
                  RPYAGITGATGAVAGAYIADITDGEDRAHFGIMSCFQSVWAGVAVAGGLIGALISLH
                  APVLAALVNGNLILGCFLMQESHKGERPMPLEAFNVSSFRNARGMTIYAALMTV
                  FFIQJLVGQVPALMNVIFGEDRFRMSATFISLAVFGILHLAOLFQVTPATKRGE
                  KOATIGMAADALGVYLLAFATRGWMAFFIMILIASGIGMLQAMLSROYDDHOG
                  OLOGSLAALTSLSITGPIIVTAIVTASSTVNGLAMIVGALLVLCPLRKAMSR
                  ATST"
misc_recomb      complement(4013..4126)
                  /note="Tn3 res site"
                  /organism="synthetic construct"
                  4161..4194
                  /note="loxP site"
                  /organism="synthetic construct"
                  <4206..>4295
                  /note="from Influenza virus"
                  /codon_start=1
                  /transl_table=11
                  /product="hemagglutinin triple-epitope"
                  /protein_id="AAB01176.1"
                  /db_xref="GI:1335941"
                  /transl_table="YPYDVPDAGYPYDVPDAGSYPYDVPDYA"
                  4305..4321
repeat_unit      /rpt_type="terminal
                  1014 C 1084 G 1091 T
BASE COUNT      1153 A
ORIGIN
Query Match      26.0%; Score 359.2; DB 12; Length 4342;

```

```

Best Local Similarity 65.7%; Pred. No. 26-69;
Matches 523; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 209 catgtcaacgaactcttaacgaagaaggccaagcacaaatlcgcagctgtcgtagaac 268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1687 CATGTCAAGAAAGCTACATATTAAGAACAGTGTCTACTACTCTAGCTGCTGACCA 1746
QY 269 gctctgcacatgaacgaagaagaacaaatctctgcctccctcgtatgctgcgac 328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1747 GCTATTTTATATCATCTACAGAAAGCAAAACAACTTGTGTCTTCACTTGGATCTGTAC 1806
QY 329 gtctagaagctctcgtgaactagacacacgcctcggagaccgacattgtctgtgaagc 388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1807 CACCAAGAAATTTACTGAGTACTGAGTACTGAGTACTGAGTACTGAGTACTGAGTACT 1866
QY 389 acctgtgcacatactgaacgactgcacacacgaagaacgaacgcctgcgcgcgcgcgc 448
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1867 ACATGTGATATCTTGTGATGATTTTTCATGAGAGGACACGTTAAGCCCTTAAAGCAAT 1926
QY 449 tgcgctaaagcaacactcatcatctcgaagaccgcaagctcgcgtgacattgcgaacac 508
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1927 ATCCGCCAAGTACAAATTTTACTCTTCCAGACACAAATTTTCTGACATTGTTAATAC 1986
QY 509 ggttaagctgcagctactcctcgcgcgtgtacacgtatcgcgagctgcgagatataccaa 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1987 AGTCAAAATTTGACGACTCTGCGGGGTGTATACAGAAATAGCAGAAATGGCAGACATTACGAA 2046
QY 569 tgcacacgcgcgcacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 628
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2047 TGCACACGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2106
QY 629 ctacacaggaacccgaagggtctgtcgtatgtcgtcgaagacacacacacacacacacac 688
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2107 AACAAAGGAACCTAGAGGCGCTTTTGTATGTATACAGAAATGTCTATGCAAGGCGCTCCCTAGC 2166
QY 689 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 748
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2167 TACTGAGATATATTAACAGGCTGACTGTGTGACATTCGAAAGACGCAAAAGATTGTGTAT 2226
QY 749 cgggttcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 808
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2227 CGGCTTTATTTGCTCAAAAGACATGGGTGGAAGACATGAGTTTACGATTGGTTATATAT 2286
QY 809 gaccgcgggggtctgcgcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 868
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2287 GACACCCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2346
QY 869 ggaatgagtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 928
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2347 GGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2406
QY 929 ggaagaagaccccaagctcgaagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 988
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2407 GGAAGGCGATGTGATAGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGGATGAGG 2466
QY 989 ggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1004
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2467 GAGAAGATGCGGCCAG 2482

```

RESULT 14

```

LOCUS      SCU54829      6107 bp      DNA
DEFINITION Synthetic E. coli Tn3-derived transposon OMP decarboxylase (URA3),
            and tetracycline efflux protein (tet) genes, complete cds, and
            beta-galactosidase (lacZ) gene, partial cds.
ACCESSION  U54829
VERSION    U54829.1  GI:1335931
KEYWORDS
SOURCE     synthetic construct.
ORGANISM   synthetic construct
REFERENCE  1 (bases 1 to 6107)

```


Db 3932 TACTGGAATATATCTAAGGCTACTGTTGACATTGCCAGACGACAAAGATTGTTAT 3991

Qy 749 cgggltcaatcgacgacgctgacatgagtgagcgacgacgcttgaatgacatcat 808

Db 3992 CGGCTTATTTCTCAAGAGCATGGTGGAGAGATGAAGGTTACGATTGGTTGATTAT 4051

Qy 809 gacccggggggttggctgagcagcaaaagagagagcgccctgggacgacgacgagtg 868

Db 4052 GACACCCGGGTGTGGCTTGAATGACAAAGGAGCGCATTTGGTTCACAGTATAGAACCGT 4111

Qy 869 ggaagagtcgcacgacgagtcaccgattgacatcgttggacgagggccttgcacaa 928

Db 4112 GGATGATGTGCTCTCTACAGATCTGACATTATATTTGTTGGAAGAGGACTATTGGCAA 4171

Qy 929 gggagagacacccaaagtcgaggggtgcccgctaccgacgaagcgcggttggagagcctt 988

Db 4172 GGGAGGAGATGCTTAAGGTAGGCGTGAACGTTACAGAAAGCAGCGCTGGAGAGCATTTT 4231

Qy 989 ggcgcgtatggcgag 1004

Db 4232 GAGAGATGGCGCCAG 4247

RESULT 15
SCU54828 6119 bp DNA SYN 21-DEC-1999

LOCUS SCU54828 2 Synthetic E. coli Tn3-derived transposon OMP decarboxylase (URA3) and tetracycline efflux protein (tet) genes, complete cds, and beta-galactosidase gene (lacZ), partial cds.

DEFINITION beta-galactosidase gene (lacZ), partial cds.

ACCESSION US4828.1 GI:1335926

VERSION US4828.1

KEYWORDS synthetic construct.

SOURCE synthetic construct.

ORGANISM 1 (bases 1 to 6119)

REFERENCE 1 (bases 1 to 6119)

AUTHORS Ross-Macdonald, P., Sheehan, A., Roeder, G.S. and Snyder, M.

JOURNAL A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae*

MEDLINE Proc. Natl. Acad. Sci. U.S.A. 94 (1), 190-195 (1997)

PUBMED 97144417

REFERENCE 2 (bases 1 to 6119)

AUTHORS Ross-Macdonald, P.B.

JOURNAL Direct Substitution

TITLE Submitted (12-Apr-1996) Petra B. Ross-Macdonald, Biology, Yale University, PO Box 208103, New Haven, CT 06520-8103, USA

FEATURES

source 1..6119

location/Qualifiers

1..6119

/organism="synthetic construct"

/db_xref="taxon:32630"

repeat_unit 1..38

/rpt_type="terminal"

misc_recomb 93..126

/note="loxR site"

gene 135..3188

/organism="synthetic construct"

/gene="lacZ"

<135..3188

/gene="lacZ"

/codon_start=1

/transl_table=11

/product="beta-galactosidase"

/protein_id="AAB01164.1"

/db_xref="GI:1335927"

/translation="DPVVLQRRHDMENPGVTOLNRLAAHPFASMRNSFEARTDPSOO
LRLNGEWFAPFAPPAEVPRESMLCEDLPADDTVVPNMGMHCYDAIYNNYPTP
VNPPEPTEPPTGCSLTNPDESILQCGRIITFDYNSAFHLMCGNRWYGGDSR
LSPERDLSAPLRAGENRLAAVYLRKSDSTLEDDMMHSGIFRVSLSLHKPTQISD
FVATRFNDPSRAVLEAEVQGLRDLRYLVTVYSLWGETQVSGAPFGEIIDER
GGYADRVTLRLVNEPKLMSAEIPNLVRAVLELHTADGTLLEACNCADGCFREVIENG
LLNLGDPULLRGVNRHEHNPJLHGOVMQEDMVDIILMKONNEFAVYCSPPNPLP
YTLCDRYGLYVDEANIEITHGVPMNRLTLDPRMLPAASEVETRMVQDRNHPVILII

SLNDSGCHGNHNDALRYMIVKSPVPEYEGGADTATDIIICPMYARVEDQPPPA
VFKWSIKKKMLSPGETRPLILCEYVHAAGNSLAGPAXKWOAFROYPPLOCGFVMDWD
OSLITVDNGNWSA YGGDEQDTPDRDPCNNGLYFADRTFRPALTEKHQOOFQPR
LSGQITVETSETLFHNSNDELIMMYAIDGKPLASGEVPLDVAQGLLELPLP
ESAGOLMLVTRVQPNATWSEVAGHISAMQOMRLAENLSVPLPASHAIPHLTTSMD
FCIELGNKRWQFNQSGFLSQMIGDKQLTLPADQFTRAPLNDIGVSEATRIDEN
MAVERMKAAHQAEALLOCTADTLADAVLITTAHAMQHOGKTLFISRTYRIDSG
QMALIVDEVASDTPPARIGLNCLOAVRVMNLGLGPOENTPDLTACPRMPL
PLSDVTPYVPFSENGLCRGTRRLNYSFQHRQDPQPNISYTSQOQLMETSRLHLIA
PEGTWLNIDGFIPIGIGDSDMSPSVSAEFLSAGRYHQLWVCKR"

3465..4268

/gene="URA3"

/gene="URA3"

3465..4268

/note="ura3 from *S. cerevisiae*"

/codon_start=1

/transl_table=11

/product="OMP decarboxylase"

/protein_id="AAB01165.1"

/db_xref="GI:1335928"

/translation="MSKATYKERRAATHPSPVAKLFINIMEKOTNLCLASDVRTTEL
LEIVPALPKICLKTHYDIITDPSMESTVYPLAKLSKTNFLPEDRKFDIGNTKL
LOYSAGVTRIAEMADITNAHGVGGLVSLGQAEEVTEPRGLMLAELSCKSLA
TGEYTKGTVDIAKSDKDEYIGFIAQRDMGDEGYDMLIMTPGVGLDDKDALQOYR
TVDDVSTGSDIIIIYVGRGLFAKGRDAKEGGERYRKAGMEAYLRGCGON"

4461..5651

/gene="tet"

4461..5651

/gene="tet"

/note="tet"

/codon_start=1

/transl_table=11

/product="tetracycline efflux protein"

/protein_id="AAB01166.1"

/db_xref="GI:1335929"

/translation="MKSNNALITVLITGTVTLDAVIGLWMPYLPGLRLDIYHSDSIASH
YGVLLATYALMQLCAPYLALSDRFGRFPLVLAISLGATIDVIMATTPVLMILVAG
RIVAGITGATGAVAGAYIADITDGEDGAEHREGLMASCFGVMVGPAGGLGASIH
APFLAAVNLNLNLGCFLOESHGKGERMPPLRANPVSPFMARGMTIVAAIMV
FEIMOLVGQVPAALAVITEGDFRFSWTGICLSAVGILHALAQAQVTPARKRRE
KQATACGAAADALCYVILATFRTRMAFPIMILASGICGAPALQAMLSQVDDHOG
OLQSLALVLSITGTLVLTALYASASTWGLVAGAAVLYCALVLCALPARGAMSR
ATST"

5790..5903

/note="tn3 res site"

/organism="synthetic construct"

5938..5971

/note="loxP site"

/organism="synthetic construct"

<5983..>6072

/note="from influenza virus"

/codon_start=1

/transl_table=11

/product="hemagglutinin triple-epitope"

/protein_id="AAB01167.1"

/db_xref="GI:1335930"

/translation="YPYDVPDYAGYPYDVPDYAGSYPYDVPDYA"

6082..6119

repeat_unit 1..38

/rpt_type="terminal"

BASE COUNT 1445 a 1563 c 1666 g 1445 t

ORIGIN

Query Match 26.0% Score 359.2; DB 12; Length 6119;
Best Local Similarity 65.7% Pred. No. 1.9e-69;
Matches 523; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

Qy 209 catgtcaacgaatacttcagcagaagggccaaagcacaacatcgcagctgtcaga 268

Db 3464 CATGTGGAAGCTACATATTAAGAAAGCGTGTACATCATCTGCTCTGTGCGCAA 3523

Qy 269 gcttcgcatcgtatgacgagaagaacaacatcctcgtcgtccctcgtatgtcgagac 328

